



THÈSE

En vue de l'obtention du DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par l'Université Toulouse 3 - Paul Sabatier

Présentée et soutenue par

Agathe MALBEC

Le 17 décembre 2018

Contrôle épigénétique de la biologie des lymphocytes T CD4

Ecole doctorale : **BSB - Biologie, Santé, Biotechnologies**

Spécialité : **IMMUNOLOGIE**

Unité de recherche :

CPTP - Centre de Physiopathologie de Toulouse Purpan

Thèse dirigée par

Joost van MEERWIJK et Olivier JOFFRE

Jury

Pr. Jonathan WEITZMAN, Rapporteur

Dr. Lennart MARS, Rapporteur

Dr. Salvatore SPICUGLIA, Rapporteur

Pr. Maha AYYOUB, Présidente du Jury

Pr. Joost van MEERWIJK, Directeur de thèse

Dr. Olivier JOFFRE, Co-directeur de thèse

A la mémoire de

Paola ROMAGNOLI

Remerciements

Je remercie infiniment le Professeur Ayyoub d'avoir accepté de présider le jury de ma soutenance de thèse et le Professeur Weitzman et les Docteurs Mars et Spicuglia pour leur implication dans leur rôle de rapporteur, c'est un honneur pour moi.

J'adresse toute ma gratitude au Professeur van Meerwijk et aux Docteurs Romagnoli et Joffre, pour m'avoir accueillie au sein de leur équipe et m'avoir communiqué leur passion de l'immunologie.

A mes directeurs de thèse, le Docteur Romagnoli et le Professeur van Meerwijk.

Paola vous m'avez permis, en nous accordant votre HDR, de dépasser les barrières administratives de l'école doctorale et d'ouvrir ce chapitre complexe mais si enrichissant de ma vie. Je n'oublierai jamais ni votre enthousiasme contagieux, ni votre sourire chaleureux qui apportait un peu de douceur. Votre disparition a laissé un grand vide entre les murs de ce laboratoire.

Joost, vous m'avez permis avec bienveillance de clore ce travail. J'ai bien retenue la leçon : les bonnes réponses sont « les chimères mixtes hématopoïétiques », « le single-cell RNA-seq » et le chocolat (même si gâteau fonctionne aussi).

A mon co-directeur de thèse qui n'a de « co-» que la terminologie administrative, le Dr. Olivier Joffre, pour son accueil lors de mon stage de M2 et pour m'avoir permis par la suite de travailler sur ce magnifique projet de thèse de l'étude de la régulation épigénétique des lymphocytes T. Je n'oublierai ni ton exigence scientifique ni ton dynamisme. Ta rigueur méthodologique restera un exemple pour moi.

Aux membres de mon comité de suivi de thèse, le Professeur Hudrisier, les Docteurs Cammas et Malnou, et ma marraine de thèse le Docteur Laffont-Pradines, pour leurs conseils et leurs encouragements.

A la région Midi-Pyrénées et à la FRM pour le financement de ma thèse.

Un grand merci aux membres du groupe épigénétique :

Le Docteur Adoue qui m'a formée au monde minutieux et complexe de la biologie moléculaire. Fournisseuse de café en chef, tu as toujours su m'encourager et me rassurer. Véro, tu es un modèle d'empathie et de douceur au service de ton équipe et des exigences du métier de chercheur... Mais je n'oublierai pas non plus ton espièglerie (certains diraient même perfidie).

Joanna, pour le soutien apporté au cours de ces quatre années. Je n'oublierai pas nos heures passées à refaire le monde, sur le parking nord après la journée de travail, ni toutes

tes petites (GRANDES) attentions quotidiennes ainsi que ton sens infailible de l'organisation (comment vais-je faire sans toi ?).

Bénédicte Binet pour sa participation au projet SETDB1.

Un grand merci également aux membres passés et présents de l'équipe JvM qui m'ont soutenue mais aussi supportée : Geneviève, Amélie, Karine, et tout particulièrement Julie N., Katy et Nikki, pour la fraîcheur qu'elles ont apporté dans le laboratoire. Une pensée toute particulière à Ariel pour sa gentillesse et à Claire pour son soutien précieux avec le phénotypage et le génotypage.

A Mehdi (alias Mehdichouuu et maintenant Docteur Attia), pour son soutien et son sens de la pédagogie. C'est grâce à toi que je maîtrise les calculs pour effectuer une expérience de rejet de greffe de moelle osseuse. Dès que j'avais un problème, tu voguais dans ma direction pour venir à la rescousse. On ne peut rêver meilleur matelot au sein d'un équipage. Merci également de m'avoir accompagnée avec patience dans la découverte de l'expérimentation animale malgré ma réticence dans ce domaine.

A Cécile pour son enthousiasme et son humour. Depuis que je te connais, je réfléchis à deux fois avant d'utiliser certains mots. Merci aussi pour les clémentines et nos discussions.

A Sarah (alias Sarouman ou petit chat selon l'humeur du jour) pour son dynamisme et son efficacité. Tu es l'artiste du génotypage, toujours prête à donner un coup de main, et experte dans l'organisation des événements pour souder le groupe.

A Julie (alias Ju ou Docteur Darrigues), mon amie de galère depuis le M2, pour ses encouragements et son soutien au cours de ces années. Quelle chance nous avons eu d'atterrir dans la même équipe ! J'ai acquis beaucoup de rigueur à tes cotés. Merci pour ton aide à l'animalerie et pour la relecture de ma thèse. Que des bons moments passés ensemble, des dimanches roller aux multiples soirées... nous continuerons à Lisbonne ou ailleurs.

A Jérémy, pour son aide dès potron-minet à l'animalerie, sa bonne humeur et son espièglerie. Merci d'avoir été mon camarade d'escalade et de patinoire, toujours partant pour une sortie. Me tiendras-tu toujours au jus des potins du CPTP ?

A toutes les équipes du CPTP qui m'ont aidée au fil de ces années. En particulier à l'équipe JCG, et surtout à Magali, Mélanie, Eve, Marion, Nicolas et Charly que j'ai eu le plaisir de côtoyer. Mais aussi à Anna Salvioni, ma copine d'escalade toujours attentive et compréhensive, et à Manuel Lebourrier, joyeux luron et pipelette du bureau, pour son soutien à chaque coup dur et pour m'avoir évité de multiples hypoglycémies.

Aux membres de l'équipe JPS qui m'ont épaulée, dépannée et avec qui j'ai eu le plaisir de partager quelques bières. Merci à Katy, j'ai adoré travailler avec toi, toujours souriante même quand les journées sont longues et les manip complexes ; à Benjamin pour ses

conseils avisés et sa pédagogie ; à Michel pour son humour ; à Cécile Goursat, travailleuse acharnée toujours prête à s'amuser une fois le travail fini. Enfin merci à Nicolas. Je ne compte plus le nombre de fois où tu m'as encouragée et apporté ton soutien au cours de ces cinq années, toujours partant pour un Friday drink, une soirée, ou une virée à la salle d'escalade.

A tous les membres du plateau de cytométrie, qui m'ont supportée jusqu'à dans les jours de grand stress et à l'équipe de la zootechnie de Purpan.

Je n'oublierai pas Francine Chauvier et sa bonne humeur, Caroline Merlo et son sourire et la disponibilité doublée d'efficacité de Dominique Rozet.

A mes amis du M2 Alice, Delphine, Mélanie, Julie, Michael et Pierre-Paul et à Núria et Jay qui ont toujours répondu présents et n'ont cessé de m'encourager!

A ma team fac : Célia, Marie et Orianne pour leur confiance, leurs encouragements et leur amitié. Vous avez compris l'investissement que me demandait cette thèse et ne m'en avez jamais tenu rigueur. Bientôt dix ans que nous nous connaissons et toujours soudées malgré la distance et le temps qui file!

A mes proches,

A Majd, pour son positivisme et son calme à toute épreuve qui ont su enlever un grand poids de mes épaules (le reggae a du aider).

A ma grand-mère, pour avoir toujours veillé aux plaisirs de mes papilles, (ses petits plats ont fait des envieux au CPTP) et toujours prêté une oreille attentive à mes tracas.

Une pensée et toute mon affection à mes grands-parents disparus.

A ma sœur, Marine. C'est à mon tour de te remercier pour tes incantations magiques, tes sauvetages téléphoniques, et surtout pour ton soutien indéfectible dans la lutte antimoustique... on les vaincra un jour.

A mes parents, une thèse ne serait pas suffisante pour vous remercier à la hauteur de tout ce que vous avez fait pour moi. Corriger les fautes de ce mémoire n'en représente qu'une infime partie. Merci pour votre soutien inconditionnel !

TABLE DES MATIERES

RESUME	5
ABSTRACT	6
LISTE DES FIGURES	7
LISTE DES TABLEAUX	8
PRINCIPALES ABREVIATIONS	9
INTRODUCTION	17
Chapitre I : Les lymphocytes T CD4	17
I. Développement intrathymique des lymphocytes T CD4	17
II. Homéostasie des lymphocytes T CD4 en périphérie	23
A. Maturation, trafic et survie des lymphocytes T CD4 naïfs	23
B. Mécanismes de tolérance périphérique	26
III. Réponses des lymphocytes T CD4 en périphérie	28
A. Les cellules dendritiques : un lien entre immunité innée et adaptative	28
B. L'activation et la différenciation des LT CD4 nécessitent 3 signaux	29
1. Le signal TCR	31
2. Le signal de co-stimulation	34
3. Le signal polarisant	35
C. Hétérogénéité et plasticité des lignages T helpers	39
1. Hétérogénéité des lignages T CD4	39
a. Les lymphocytes de type Th1	39
b. Les lymphocytes de type Th2	40
c. Les lymphocytes de type Th17	42
d. Les lymphocytes T follicular helper (T _{FH})	43
2. Du paradigme Th1/2... au continuum de différenciation	45
Chapitre II : Régulation de l'expression des gènes	51
I. Organisation de la chromatine	51
II. Eléments <i>cis</i> -régulateurs de l'expression génique	54
III. Régulation épigénétique	57
A. Historique et concept	57
B. Méthylation de l'ADN	60
1. ADN méthyl transférases (DNMT)	61
2. Rôle de la méthylation de l'ADN	63
3. La méthylation de l'ADN est réversible	64
C. Modification post-traductionnelles des histones	66
1. Codes histones	66
2. Acétylation des histones	67
3. Méthylation des histones	69
a. Méthylation d'H3K4	70

b. Méthylation d'H3K27	71
D. Méthylation d'H3K9	72
1. Mono-, di- ou tri-méthylation d'H3K9	72
2. Localisation et fonctions d'H3K9me3	72
a. Composant de l'hétérochromatine	72
b. Répression des promoteurs	73
c. Répression des enhancers	74
d. Répression des éléments transposables	75
e. Autres fonctions régulatrices	76
3. H3K9 Lysine méthyl transférase	77
4. H3K9me3 readers: L'interactome HP1	79
5. H3K9me3 déméthylases	80
Chapitre III : Régulation épigénétique de la biologie des lymphocytes T CD4	83
I. Régulation épigénétique du développement intrathymique des lymphocytes T CD4	83
II. Régulation épigénétique des lymphocytes T CD4 en périphérie	89
A. Régulation épigénétique des fonctions effectrices	91
1. Remodelage de la chromatine par les facteurs pionniers	91
a. Activation des enhancers par les STAT	91
b. Déplacement des nucléosomes et réorganisation spatiale de la chromatine	92
c. Exclusion du complexe Polycomb	94
2. La dé-méthylation de l'ADN	94
B. Répression épigénétique des lignages alternatifs	97
1. Méthylation de l'ADN	97
2. La voie Polycomb – H3K27me3	97
3. Relocalisation des gènes	98
C. Rôle de l'interactome d'H3K9me3 dans la biologie des LT	99
CONTRIBUTIONS	101
RESULTATS	103
SETDB1 controls T helper cell lineage integrity by repressing endogenous retroviruses	103
I. SUMMARY	103
II. INTRODUCTION	104
III. RESULTS	106
A. Th1 priming is enhanced in the absence of SETDB1	106
B. Impaired acquisition of the Th2 phenotype by SETDB1-deficient cells	108
C. SETDB1 is required for stable Th2 cell commitment	109
D. SETDB1-dependent H3K9 trimethylation at a subset of ERVs	110
E. SETDB1-deficiency upregulates ERVs and their neighboring genes	111
F. H3K9me3+ ERVs mark the Th1 enhancer landscape in Th2 cells.	112
G. SETDB1-dependent H3K9 trimethylation at ERVs represses Th1-specific enhancers.	114
IV. DISCUSSION	115

V. FIGURES.....	118
MATERIEL ET METHODE.....	139
DISCUSSION ET PERSPECTIVES	155
I. Coopération des mécanismes épigénétiques dans la répression de lignages alternatifs de lymphocytes T helper	156
II. SETDB1 régule le programme Th1 <i>via</i> la déposition H3K9me3 sur un set restreint d'ERV	159
A. Propagation locale d'hétérochromatine	159
B. Evolution fonctionnelle des ERV en éléments <i>cis</i> -régulateurs	159
C. ERV module <i>cis</i> -régulateur ou DAMP	161
III. Ciblage de SETDB1 sur les ERV	163
IV. Coopération des mécanismes épigénétiques dans la répression des ERV	165
V. Intérêts thérapeutiques de la modulation de l'axe SETDB1-H3K9me3-ERV	166
REFERENCES	169

RESUME

Les lymphocytes T CD4 naïfs sont des cellules plastiques, capables de moduler finement leur programmation selon les signaux environnementaux qu'ils intègrent. Ils adaptent ainsi leur phénotype et leur fonction au type de danger lors d'une infection par un agent pathogène intracellulaire par exemple, ils acquièrent un phénotype Th1 sous l'influence de médiateurs solubles tels que l'IL-12 et l'IFN- γ . Ces signaux mobilisent un set restreint de facteurs de transcription, coordonné par Tbet, qui programment la cellule afin qu'elle induise l'élimination du danger par des mécanismes impliquant une production massive d'IFN- γ . En réponse à des allergènes ou à des parasites extracellulaires, les lymphocytes T peuvent aussi acquérir un phénotype Th2, caractérisé par l'expression du facteur de transcription Gata-3 et par la production d'IL-4, d'IL-5 et d'IL-13. Afin de garantir la stabilité des lignages, ces processus de différenciation peuvent s'accompagner d'une perte de potentialité. Contrairement aux cellules T naïves, les cellules Th1 sont par exemple incapables d'allumer le programme d'expression génique Th2 en présence d'IL-4, et les lymphocytes Th2 verrouillent le programme Th1. Si nous savons aujourd'hui que l'acquisition des fonctions effectrices, comme l'équilibre entre détermination cellulaire et plasticité, sont régulés par des mécanismes épigénétiques, la plupart des acteurs moléculaires qui contrôlent la programmation des lymphocytes T au niveau de la chromatine reste encore à identifier. Durant ma thèse, j'ai étudié le rôle de la lysine méthyltransférase SETDB1, qui catalyse la di- ou tri-méthylation de la lysine 9 de l'histone 3 (H3K9me3), dans la différenciation des lymphocytes T CD4. Il avait déjà été proposé qu'H3K9me3 ait un impact sur la programmation de ces cellules en réponse aux signaux de l'environnement, mais personne n'avait encore étudié le rôle de SETDB1 dans ces processus lorsque j'ai commencé ma thèse. A l'aide d'une lignée murine déficiente pour SETDB1 spécifiquement dans les lymphocytes T, nous avons montré *in vitro* et *in vivo* que la balance Th1/Th2 est fortement augmentée en l'absence de l'enzyme, et que cette dérégulation résulte d'une perte de répression du réseau génique Th1. Au niveau mécanistique, nos analyses moléculaires ont mis en évidence que l'absence de SETDB1 dans les lymphocytes Th2 s'accompagne d'une perte d'H3K9me3 sur un réseau restreint de rétrovirus endogènes (ERV) qui se comportent comme des modules cis-régulateurs de gènes clefs du programme Th1. Nous avons donc mis en évidence que SETDB1 garantit l'intégrité des lignages T helpers en réprimant, via la déposition d'H3K9me3, un set restreint d'ERV exaptés en enhanceurs de gènes Th1.

ABSTRACT

Upon activation, naïve CD4 T cells differentiate into distinct helper or regulatory T cell subsets depending on environmental signals received. This process relies on complex and lineage-specific gene expression programs whose dynamics and stability are regulated at the level of the chromatin. The epigenetic pathways involved, however, remain largely unknown. Here, we report that the histone methyltransferase SETDB1 critically controls the Th1 gene expression program. SETDB1-deficient naïve CD4 T cells show exacerbated Th1 priming, and when exposed to a Th1-instructive signal, SETDB1-deficient Th2 cells cross lineage boundaries and transdifferentiate into Th1 cells. Surprisingly, SETDB1 does not appear to control Th1 gene promoter activity. Instead, it deposits the repressive H3K9me3 mark at a restricted and cell-type specific set of endogenous retroviruses (ERVs) strongly associated with genes involved in immune processes. Refined bioinformatic analyses indicated that these retrotransposons either flank and repress Th1 gene *cis*-regulatory elements or behave themselves as Th1 gene enhancers. In conclusion, H3K9me3 deposition by SETDB1 ensures T cell lineage integrity by repressing a repertoire of ERVs that have been exapted into *cis*-regulatory modules to shape and control the Th1 gene network.

LISTE DES FIGURES

Figure 1: Principales sous-populations de thymocytes	18
Figure 2 : Développement intra-thymique et détermination des LT	22
Figure 3 : Circulation des LT CD4 dans les nœuds lymphatique	25
Figure 4 : Modèle d'interaction des lymphocytes T avec les cellules dendritiques et synapse immunologique	31
Figure 5 : Cascade moléculaire en aval de la signalisation du complexe TCR.....	33
Figure 6 : Modèle de polarisation des lymphocytes T CD4	38
Figure 7 : Différents signaux environnementaux instruisent la différenciation d'un large répertoire de lignages T CD4 ⁺ caractérisés par des signatures moléculaires et des fonctions distinctes.....	44
Figure 8 : Un large panel de cytokines influence la différenciation et la plasticité des LT CD4 ⁺	49
Figure 9 : Constitution du nucléosome.....	51
Figure 10 : Organisation de la chromatine	52
Figure 11 : Compaction de la chromatine.....	53
Figure 12 : Eléments <i>cis</i> -régulateurs de l'expression génique.....	56
Figure 13 : Le paysage épigénétique de Conrad Waddington	58
Figure 14 : "Chromatin writer", "reader" et "eraser"	59
Figure 15 : Mécanismes de ciblage des DNMT à l'ADN.....	62
Figure 16 : Mécanismes répresseurs de la transcription par la méthylation de l'ADN..	64
Figure 17 : Méthylation des cytosines.....	65
Figure 18 : Modifications post-traductionnelles des histones	66
Figure 19 : Méthylation des lysines.....	70
Figure 20 : Principales lysine méthyl transférases impliquées dans la méthylation d'H3K9.....	77
Figure 21 : Exemple d'interactome de SETDB1	81
Figure 22 : La différenciation des lymphocytes T CD4 naïfs s'accompagne d'un remodelage de la chromatine.....	90
Figure 23 : Modèle de déplacement des nucléosomes et réorganisation de la chromatine	93
Figure 24 : La dé-méthylation des cytosines par les protéines TET est nécessaire à l'acquisition des fonctions effectrices par les lymphocytes T helper.....	95
Figure 25 : Modèle de remodelage permissif et répressif de la chromatine.....	96
Figure 26 : Répression épigénétique du locus <i>Ifng</i> dans les lymphocytes Th2	157

Figure 27 : Impact du type de lymphocytes T CD4 infiltrant les tumeurs sur le pronostic vital des patients.....	167
---	-----

LISTE DES TABLEAUX

Tableau 1: Principaux acteurs moléculaires impliqués dans la régulation épigénétique du développement intrathymique des lymphocytes T CD4	87
Tableau 2: Rôle de l'interactome d'H3K9me3 dans la biologie des lymphocytes T CD4	100

PRINCIPALES ABREVIATIONS

5cac	5-carboxycytosine
5fC	5-formylcytosine
5hmC	5-hydroxymethylcytosines
5mC	5-methylcytosine
ADAP	Adhesion and Degranulation promoting Adaptator Protein
ADN	Acide Désoxyribonucléotide
AGO2	Argonaute-2
AIM2	Absent In Melanoma 2
AIRE	Autoimmune Regulator
AKT	Serine Thréonine Kinase 1
AML	Leucémie Myéloïde Aigüe
AP1	Activator Protein 1
APC	Antigen Presentating Cell – Cellule présentatrice d’antigène
ARN	Acide Ribonucléique
ARNInc	ARN long non codant
ASE	Anti-Silencer
ATP	Adénosine Triphosphate
Aza	5-Aza-2-deoxycytidine
BAD	Bcl-2-Associated Death promoter
BATF	B-Cell-Activating Transcription Factor
BCL	B Cell Lymphoma
BCR	B Cell Receptor
BER	Base Excision Repair
BIM	BCL-2 Interacting Mediator of cell-death
BMP	Bone Morphogenetic Protein
BRG	Brahma Related Gene
CAF-1	Chromatin Assembly Factor 1
CARMA1	CARD- containing MAGUK protein 1
CCL	C-C Motif Chemokine Ligand
CCR	C-C Motif Chemokine Receptor
CDx	Cluster de Différenciation x
CD	Domaine chromo
CDC	Cell Division Control protein
cDC	Conventional Dendritic Cell
CDX2	Caudal Type Homeobox 2
CFSE	Carboxy Fluorescein Succinimidyl Ester
cGAS	Cyclic GMP-AMP synthase
ChIP-qPCR	Chromatin ImmunoPrecipitation - quantitative Polymerase Chain Reaction
ChIP-seq	Chromatin ImmunoPrecipitation - sequencing

CLR	Récepteur Lectines de type C
CMH	Complexe Majeur d'Histocompatibilité
CpG	Di-nucléotide Cytosine - Guanine
CRAC	Calcium -Release- Activated Calcium channel
CRISPR/Cas9	Clustered Regularly Interspaced Short Palindromic Repeats / CRISPER Associated Protein 9
CSD	Chromo-Shadow Domain
cSMAC	Central Supra Molecular Activation Cluster
CTCF	CCCTC-binding Factor
cTEC	Cellules épithéliales du cortex thymique
CTLA-4	Cytotoxique T Lymphocytes Antigen 4
CXCR3	C-X-C Motif Chemokine Receptor
DAG	Diacylglycerol
DAMP	Danger-Associated Molecule Pattern
DC	Cellule dendritique
DEC205	Dendritic Cell Receptor for Endocytosis 205
DLL1	Delta like 1
DN	Double Négative
DNMT	ADN méthyltransférases
DNMT3A/B	DNA(cytosine-5)-méthyltransferase 3A/3B
DP	Double Positif
dSMAC	Distal Supra Molecular Activation Cluster
D _β	Segment Diversity du <i>Tcrb</i>
E4p	Enhancer du gène Cd4
EAE	Encéphalite Auto-immune Expérimentale
EED	Embryonic Ectoderm Development
EGR3	Early Growth Response
ELF5	E74-Like Factor 5
EOMES	Eomesodermin
ER	Endoplasmic Reticulum
ERK	Extracellular signal-Regulated Kinase
ERV	Endogenous Retrovirus
ETP	Early T cell Progenitor
EZH	Enhancer Of Zeste 2 Polycomb Repressive Complex
FcγRIIB	Fragment constant de l'IgG II b
FEZF2	Forebrain Embryonic Zinc Finger-Like Protein 2
FISH	Fluorescence in situ hybridization
FOXP3	Forkhead box P3
FRC	Cellules fibroblastiques réticulaires
G9A	Histone-Lysine N-Methyltransferase 2 ou EHMT2
GADS	GRB2-related adaptor protein 2
GATA3	GATA-Binding Factor 3
GFI1	Growth Factor Independent 1 Transcriptional Repressor

GFP	Green Fluorescent Protein
GLP	G9A Like Protein
GLUT1	Glucose Transporter Type 1
GM-CSF	Granulocyte Macrophage Colony Stimulating Factor
GRB2	Growth factor Receptor-Bound protein 2
HAT	Histones Acétyle Transférases
HC	Hétérochromatine
HDAC	Histones déacétylases
HDM	Histones déméthylases
HEL	Hen Egg Lysosyme
HEV	High Endothelial Venule
HP1	Heterochromatin protein 1
HS	DNaseI Hypersensitivity Site
HSPC	Cellules souches hématopoïétiques
IAP	Intracisternal A-type Particule
ICAM	Intracellular Adhesion Molecule
ICOS-L	Inducible T Cell Costimulator Ligand
IFI16	IFN- γ inducible protein 16
IFN	Interferon
I κ B	Inhibitor kappa-B
IL	Interleukine
IL-xR	Récepteur à l'interleukine x
InsP3	Inositol 1,4,5-triphosphate
IP3R1	Inositol 1,4,5-trisphosphate receptor type 1
IPA	Ingenuity Pathways Analysis
iPSC	Cellules souches pluripotentes induites
IRF	Interferon Response Factor
ISG	Interferon Stimulated Genes
ITAM	Immunoreceptor Tyrosine-based Activation Motif
ITCH	Itchy E3 Ubiquitin Protein Ligase Homolog
ITK	IL-2-inducible T cell Kinase
JAKs	Janus Kinases
JmjC	Jumonji domaine containing protein
JNK	Jun N-terminal kinase
J β	Segment Joint du <i>Tcrb</i>
KAISO	Zinc Finger And BTB Domain-Containing Protein 33
KAP-1	Corépresseur transcriptionnel TRIM28
kBBE	κ B Binding Element
kDA	Kilo Dalton
KDM	Lysine déméthylase
KID	KAP-1 Interaction Domain
Kif3a	Kinesin Family Protein 3A

KLF4	Kruppel-like factor 4
KMT	Lysine méthyltransférases
KO	Knock-out
KRAB	DNA-binding KRüppel-Associated Box
KZFP	KRAB domain and Zinc Finger Protein
LAT	Linker for the activation of T cells
LB	Lymphocytes B
LCK	Leukocyte C-Terminal Src Kinase
LCMV	Lymphocytic ChorioMeningitis Virus
LFA-1	Lymphocyte function-associated antigen 1
LINE	Long-Interspersed Nucleotide Element
LPS	Lipopolysaccharide
LT	Lymphocytes T
LT _N	Lymphocytes T naïfs
LTR	Long Terminal Repeat
MAF	V-maf musculoaponeurotic fibrosarcoma oncogene homolog
MALT1	Mucosa- associated lymphoid tissue lymphoma translocation protein 1
MAML1	Mastermind-Like 1
MAPK	Mitogen activated protein kinase
MBD1	Methyl-CpG-binding protein 1
MDA5	Melanoma Differentiation-Associated Protein 5
Mdc	Macrophage-Derived Chemokine
MeCP	Complexe represses transcription
MeCP2	Methyl CpG binding protein 2
MEF	Fibroblastes embryonnaire murin
MEKK	MAP/ERK kinase kinase
mESC	Cellules souches embryonnaires murines
mHSC	Cellules souches hématopoïétique murines
MKK	MAPK kinase
MLL2	Mixed-Lineage Leukemia
MLV	Murine Leukemia Virus
MMTV	Mouse Mammary Tumor Virus
MPT	Modification Post-Traductionnelle
mTEC	Cellule épithéliale thymique médullaire
mTOR	Mammalian Target Of Rapamycine
NAD	Nicotinamide adénine dinucléotide
NFAT	Nuclear factor of activated T-cells
NFκB	Nuclear factor-κB
NHEJ	Non-Homologous End-Joining
NK	Cellule Natural Killer
NL	Nœud Lymphatique

NLR	Récepteurs NOD-like
NPC	Neuronal progenitor cell
OL-II	Organes Lymphoïdes Secondaires
OX40	TNF Receptor Superfamily Member 4
PAMP	Pathogen-Associated Molecular Pattern
PcG	Polycomb
PCNA	Proliferating Cell Nuclear Antigen
PH	Pleckstrin Homology
PHD	Plant Homeodomain
PI3K	Phosphoinositide 3-Kinase
PIAS1	Protein Inhibitor of the Activated STAT1
PIP2	Phosphatidyl-Inositol bi-Phosphate
PKC θ	Protéine Kinase C θ
PLC γ 1	Phospholipase C γ 1
pMHC	Peptide bound to MHC molecules
PNAD	Peripheral Nod Adressin
PRC2	Complexe Répresseur Polycomb 2
PRDM	PR/SET Domain
iPSC	Induced Pluripotent Stem Cell
PRR	Pattern Recognition Receptor
pSAMC	Peripheral Supra Molecular Activation Cluster
PtdIns(4,5)P2	Phosphatidylinositol 4,5-bisphosphate
RAG	Recombination Activation Gène
RASGRP1	RAS Guanyl-Releasing Protein 1
RBPJ	Recombination-signal Binding Protein for immunoglobulin kappa J region
RELM	Resistin Like Molecule
RIG-1	Retinoic acid-Inducible Gene I
RING	Really Interesting New Gene
RLR	Récepteur RIG-I-like
RNA pol II	ARN polymérase II
RNA-seq	Séquencage d'ARN
ROR γ T	Retinoic acid (RA)-related orphan receptor γ thymus
RTE	Recent Thymic Emigrant
RUNX3	Runt Related Transcription Factor 3
S1P	Sphingosine 1 phosphate
S1PR1	Récepteur à la Sphingosine 1 phosphate
S4	Elément intra-génique répresseur du gène <i>Cd4</i>
S6K	SH2-domain-containing leukocyte protein of 76 k
SAM	S-Adenosyl-L-Methionine
SATB1	Special AT-rich sequence-binding protein-1
SET	SUV39 , Enhancer of zeste and the homeobox gene regulator Trithorax
SETDB1	SET Domain Bifurcated 1

SINE	Short-Interspersed Nucleotide Element
SIRP	Signal Regulatory Protein
SLP76	SH2-domain- containing leukocyte protein of 76 kDa
SMAC	Supra Molecular Activation Cluster
SMAD	Mothers Against Decapentaplegic homolog 4 (Drosophila)
SOCS	Suppressor Of Cytokine Signaling
SOS1	Son Of Sevenless homologue 1
SP	Simple positive
SRA	SET and RING Associated
STAT	Signal Transducers and Activators of transcription
SUMO-E3	Small Ubiquitin-like MOdifier proteins-E3
SUV39H1	Suppressor of variegation 3-9 homologues
SUZ	Suppressor of Zeste
Tbet	T-Box Expressed in T cell
Tbx21	T-box 21
TCF-1	T Cell Factor
TCR	T Cell receptor
TDG	Thymine DNA Glycosylase
TDT	Terminal Desoxynucleotidyl Transferase
TE	Élément transportable
TET	Ten Eleven Translocation
TFG	Transforming growth factor
T _{FH}	T follicular helper
Th	T helper
TLR	Récepteurs Toll Like
TNF	Tumor Necrosis Factor
TRIF	Toll interleukin 1 Receptor domain-containing adapter protein Inducing Interferon Beta
TRIM28	TRIMpartite Motif-containing28
TrxG	Trithorax Groupe
TSP	Thymus Settling Progenitors
TSS	Transcription Start Site
TTD	Tandem Tudor Domain
Tx	Trithorax
Ub	Ubiquitine
UHRF1	Ubiquitine-like, containing PHD and RING finger domains
VAV1	Vav guanine nucleotide exchange factor 1
VL30	Virus Like 30
V _β	Segment Variable du <i>Tcrb</i>
WDR5	WD Repeat Protein 5
ZAP70	Zeta-Chain Associated Protein Kinase, 70kD
ZF	Zinc Finger

ZFP	Zinc Finger protein
ZNF	Zinc Finger
CBL-B	Casitas B-Lineage Lymphoma Proto-Oncogene B
GRAIL	Gene Related To Anergy In Lymphocytes Protein

INTRODUCTION

Chapitre I : Les lymphocytes T CD4

Les lymphocytes T CD4 sont des cellules clé du système immunitaire adaptatif. Bien que d'origine hématopoïétique, ils ne se développent pas dans la moelle osseuse mais dans un organe spécialisé, le thymus, localisé dans le médiastin antéro-supérieur. Au cours de leur développement, ils subissent des processus de sélection extrêmement stricts qui permettent la survie des cellules utiles pour l'organisme tout en purgeant les précurseurs autospécifiques du répertoire T qui migre en périphérie. Après émigration, les lymphocytes T finissent leur maturation dans les organes lymphoïdes secondaires avant, éventuellement, de s'activer si le danger dont ils sont spécifiques menace l'intégrité de l'organisme. Dans ce chapitre, j'aborderai de façon séquentielle ces différentes facettes de la vie des lymphocytes T CD4. Après avoir décrit brièvement leur développement intra-thymique, je présenterai les mécanismes moléculaires et cellulaires qui contrôlent l'homéostasie des lymphocytes T CD4 en périphérie. Je détaillerai enfin comment ces cellules se programment et acquièrent leurs fonctions effectrices en réponse à des signaux de danger endogènes ou exogènes.

I. Développement intrathymique des lymphocytes T CD4

Le thymus est un organe lymphoïde primaire qui involue avec l'âge. Il est le siège du développement des différents lignages de lymphocytes T qui suit un processus complexe et implique de nombreux partenaires cellulaires et moléculaires. D'un point de vue anatomique, le thymus s'organise en deux régions fonctionnellement distinctes, la médulla et le cortex, séparées l'une de l'autre par une jonction cortico-médullaire accueillant des vaisseaux sanguins. C'est au niveau de cette jonction qu'affluent, via la circulation sanguine, des progéniteurs hématopoïétiques pluripotents provenant du foie chez le fœtus et de la moelle osseuse chez l'enfant et l'adulte. Une fois dans le thymus, ce pool hétérogène de progéniteurs TSP (*Thymus Settling Progenitors*) se différencie vers le stade d'ETP (*Early T cell Progenitor*) et migre dans le cortex. Les ETPs sont des cellules immatures n'exprimant ni le corécepteur CD4 ni le corécepteur CD8. Elles sont ainsi qualifiées de « double-négatives » (DN). En réponse aux signaux délivrés par les cellules du stroma thymique, ces progéniteurs vont peu à peu perdre une partie de leur pluripotence et s'engager dans le lignage T. Différentes populations de thymocytes DN peuvent être identifiées selon leur niveau d'expression de CD25, la chaîne α

du récepteur à l'IL-2, et de la molécule d'adhésion CD44 (**Figure 1**). On distingue ainsi les cellules DN1 (CD44+CD25-), DN2 (CD44+CD25+), DN3 (CD44-CD25+) et DN4 (CD44-CD25-). Les cellules DN4 acquièrent l'expression des corécepteurs CD4 et CD8 et sont alors définies comme « double-positives ». Elles « choisissent » enfin l'un des deux principaux lignages et deviennent ainsi CD4 ou CD8 simple positives (SP) (1, 2).

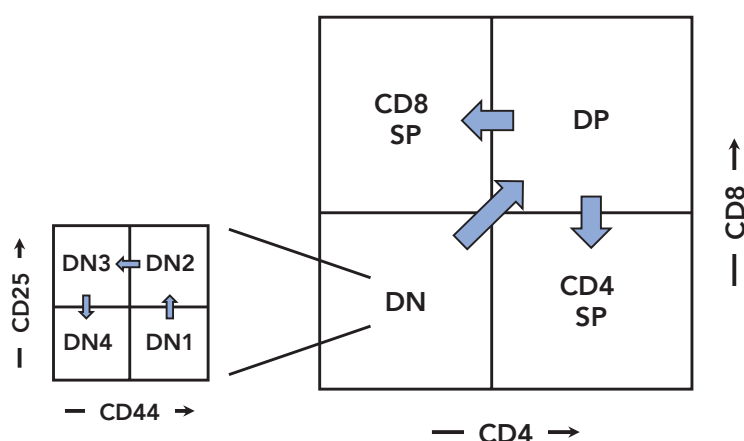


Figure 1: Principales sous-populations de thymocytes

La population de thymocytes qualifiée de « double-négative » (DN), car n'exprimant aucun des co-récepteurs CD4 et CD8, peut être subdivisée en quatre sous-populations selon le niveau de maturation et sur la base de l'expression de la molécule d'adhésion CD44 et de CD25, la chaîne α du récepteur à l'IL-2. On distingue ainsi les cellules DN1 (CD44+CD25-), DN2 (CD44+CD25+), DN3 (CD44-CD25+) et DN4 (CD44-CD25-). Les thymocytes DN4 prolifèrent activement et se différencient en cellules « double-positives » (DP). Les cellules DP qui expriment un récepteur à l'antigène fonctionnel s'engagent ensuite, sans doute de façon stochastique, dans le lignage CD4 ou CD8 et deviennent alors « simple positives » (SP).

Les premières étapes du développement intrathymique des lymphocytes T sont caractérisées par une prolifération intense des thymocytes DN1 à DN2b. Ces cellules sont multipotentes puisqu'elles ne sont pas totalement déterminées dans le lignage lymphocytaire T et peuvent encore se différencier vers des lignages myéloïdes tels que celui des cellules dendritiques (3). Leur prolifération et leur survie sont sous le contrôle de signaux exogènes délivrés par l'épithélium thymique cortical sous la forme notamment de facteurs de croissance comme le *Stem Cell Factor*. Ces cellules stromales expriment aussi des molécules de surface, telles que Delta Like 1 (DLL1) et DLL4, qui après interaction avec le récepteur Notch à la surface des thymocytes mobilisent les facteurs de transcription TCF-1, GATA3 et Bcl11b. Ces molécules engagent alors le précurseur dans le lignage lymphocytaire T, et induisent *de facto*

une perte de multipotence (4–6). L'importance de ces facteurs dans le développement des lymphocytes T a notamment été soulignée par le groupe d'E. Rothenberg. Ils ont montré que la déplétion de *Bcl11b* dans des cellules souches hématopoïétiques bloque la différenciation des thymocytes aux stades DN2b/DN3 au profit d'une différenciation des thymocytes DN1 et DN2 vers le lignage myéloïde (7).

Aux stades cellulaires DN2b et DN3, les thymocytes finissent de se déterminer dans le lignage lymphocytaire T. Leur prolifération diminue et *Rag1* et *Rag2*, deux gènes essentiels au réarrangement des segments de gènes codant pour les chaînes alpha et beta du récepteur à l'antigène (*T cell receptor*, TCR), sont exprimés. Ils codent pour des protéines recombinases homonymes qui constituent les sous-unités catalytiques des complexes RAG. Il a été proposé que les gènes *Rag* soient issus de transposons, des éléments génétiques capables de se déplacer de manière autonome dans le génome. L'intégration de génome viral codant pour une activité recombinase pourrait donc être un événement clef dans la mise en place du système immunitaire adaptatif tel que nous le connaissons aujourd'hui chez les vertébrés à mâchoire (8, 9).

Dans toutes les cellules de l'organisme à l'exception des cellules T, les gènes codant pour la chaîne alpha et bêta du TCR sont non fonctionnels. Les multiples segments de gènes V (*Variable*), D (*Diversity*) et J (*Joint*) du locus *Tcrb*, et V et J du locus *Tcra*, doivent en effet être réarrangés de façon séquentielle et aléatoire dans les thymocytes en cours de développement avant de pouvoir coder les deux sous-unités de l'hétérodimère. Ces événements de recombinaison sont coordonnés par RAG1 et RAG2, les cassures d'ADN générées étant quant à elles réparées par le complexe NHEJ (*Non-Homologous End-Joining*). Cette voie de réparation, ubiquitaire, fait notamment intervenir la Terminal desoxynucleotidyl Transferase (TdT) qui peut ajouter aléatoirement quelques nucléotides N (*Nontemplated*) au niveau des cassures, créant ainsi une grande diversité jonctionnelle en plus de la diversité combinatoire (il existe respectivement 52, 2 et 13 segments V, D et J fonctionnels dans le locus *Tcrb*). L'expression de RAG1 et RAG2 est essentielle au développement des lymphocytes. En effet des études ont montré chez l'homme comme chez la souris que des gènes *Rag* non fonctionnels associés à une immunodéficience sévère notamment caractérisée par une absence de cellules T et B matures (10, 11). Après réarrangement des segments VDJ du locus *Tcrb*, un complexe pré-TCR est exprimé à la surface des cellules. Il est constitué d'un hétérodimère dans lequel la chaîne β du TCR s'associe à la pseudo-chaîne invariante α ($pT\alpha$) et aux complexes de signalisation CD3 qui existent sous la forme d'hétérodimères $CD3_{\gamma\epsilon}$ et $CD3_{\delta\epsilon}$ et d'homodimères $CD3_{\zeta\zeta}$ (12, 13). La signalisation du complexe pré-TCR, qui témoigne d'un réarrangement fonctionnel du locus *Tcrb*, délivre un signal de survie aux thymocytes et

leur permet de progresser dans leur développement jusqu'au stade DN4. A ce stade, les thymocytes perdent leur dépendance à la signalisation Notch pour leur survie et sont pleinement engagés dans la lignée lymphoïde (14).

Une fois la sélection β passée, les thymocytes DN3 et DN4 prolifèrent intensément et progressent jusqu'au stade double positif (DP). Ils arrêtent alors de cycler, internalisent le complexe pré-TCR, et ré-expriment *Rag2* ainsi que toute la machinerie de recombinaison afin de réarranger les segments $V\alpha$ et $J\alpha$ qui codent pour la région variable de la chaîne α du TCR. Si les événements de recombinaison aboutissent à un réarrangement fonctionnel du locus *Tcra*, la chaîne alpha s'associe à la chaîne bêta et l'hétérodimère est exprimé à la surface associé aux modules de signalisation CD3. Les deux chaînes du complexe $TCR\alpha\beta$ étant générées et appariées de façon aléatoire, seulement 15% du répertoire de thymocytes DP exprime un récepteur à l'antigène capable de reconnaître le CMH du soi à la surface des cellules stromales (15). Afin d'éliminer les cellules inutiles à l'organisme, les thymocytes subissent alors une étape de sélection dite « positive ». Seules les cellules DP dont le récepteur à l'antigène reconnaît le CMH du soi à la surface des cellules épithéliales du cortex thymique (cTEC) intègrent un signal de survie et poursuivent leur développement vers les stades SP (16, 17). Les autres meurent « par négligence ». Les mécanismes qui régulent l'engagement dans le lignage CD4SP *versus* CD8SP ne sont pas complètement élucidés à ce jour. Il a initialement été proposé que la différenciation des thymocytes DP vers le lignage CD8SP puisse être induite par l'interaction de leur TCR avec un complexe CMH-I-peptide, alors qu'une interaction avec un complexe CMH-II-peptide les engagerait vers le lignage CD4 SP. Mais ce modèle « instructif » a été remis en question par d'autres travaux défendant un modèle « stochastique » dans lequel le thymocyte s'engagerait dans l'un ou l'autre des lignages de façon aléatoire, indépendamment de la restriction de son TCR pour les molécules du CMH. Dans des souris transgéniques dans lesquelles toutes les cellules T expriment un TCR restreint pour une molécule du CMH I, des thymocytes s'engagent en effet quand même dans le lignage CD4. Les mécanismes supportant la différenciation de thymocytes DP en cellules SP CD4 ou CD8 restent donc sujet à débat (18) (**Figure 2**).

Au stade SP, les thymocytes « uprégulent » le récepteur CCR7 ce qui leur permet de migrer dans la médulla thymique pour y poursuivre leur maturation en suivant le gradient de chimiokines CCL19 et CCL21 (19). Les processus de réarrangement des chaînes alpha et beta du TCR étant aléatoires, une partie du répertoire est auto-spécifique. A ce stade, de nombreux précurseurs expriment en effet un récepteur à l'antigène reconnaissant un complexe CMH-peptide du soi et sont donc susceptibles d'induire une réaction auto-immune s'ils deviennent matures et migrent en périphérie. Une étape de sélection « négative » permet d'éliminer ces

thymocytes potentiellement dangereux pour l'organisme. Elle implique les différentes sous-populations de cellules dendritiques présentes dans le thymus et les cellules épithéliales thymiques médullaires (mTECs). Ces dernières expriment les facteurs de transcription AIRE et FEZF2 qui leur permettent d'exprimer de façon ectopique des antigènes tissulaires, et de projeter ainsi à leur surface le soi immunologique de l'individu (20). Les thymocytes SP peuvent donc tester la spécificité de leur TCR contre un large répertoire de complexes CMH-peptide. Pour les cellules engagées dans le lignage CD4, ces interactions peuvent avoir trois conséquences sur le devenir du thymocyte en fonction de l'affinité entre leur récepteur à l'antigène et les complexes exposés à la surface des mTECs et des DC. Ceux ayant une haute affinité pour leur ligand, c'est-à-dire étant auto-spécifiques, sont inactivés fonctionnellement (on parle d'état « anergique ») ou éliminés physiquement par apoptose. A l'opposée, les précurseurs ayant un TCR de faible affinité survivent et se différencient en cellule T conventionnelles. Enfin, ceux reconnaissant des antigènes du soi avec une affinité intermédiaire entre ces deux situations sont préférentiellement déviés vers le lignage régulateur. Ils acquièrent alors l'expression du facteur de transcription FOXP3, et après migration dans les organes lymphoïdes secondaires, auront pour principale fonction de contrôler de façon qualitative et quantitative les réponses immunitaires. Ils ont notamment un rôle clef dans la répression du répertoire lymphocytaire T auto-spécifique qui aurait pu échapper à ces processus de sélection (21).

Sur 100 thymocytes DN qui entament leur développement dans le thymus, seuls 3 à 5 franchissent toutes ces étapes de sélection. Ce processus de développement est donc extrêmement coûteux d'un point de vue énergétique pour l'organisme. Il a néanmoins été conservé au cours de l'évolution car il permet de générer un répertoire de cellules T capables virtuellement de reconnaître n'importe quel type de danger tout en éliminant les précurseurs qui pourraient être délétères pour l'organisme. Afin de préserver l'intégrité de l'individu, ces cellules migrent alors dans les organes lymphoïdes secondaires. Dans la suite de cette thèse je me concentrerai uniquement sur la biologie des LT CD4.

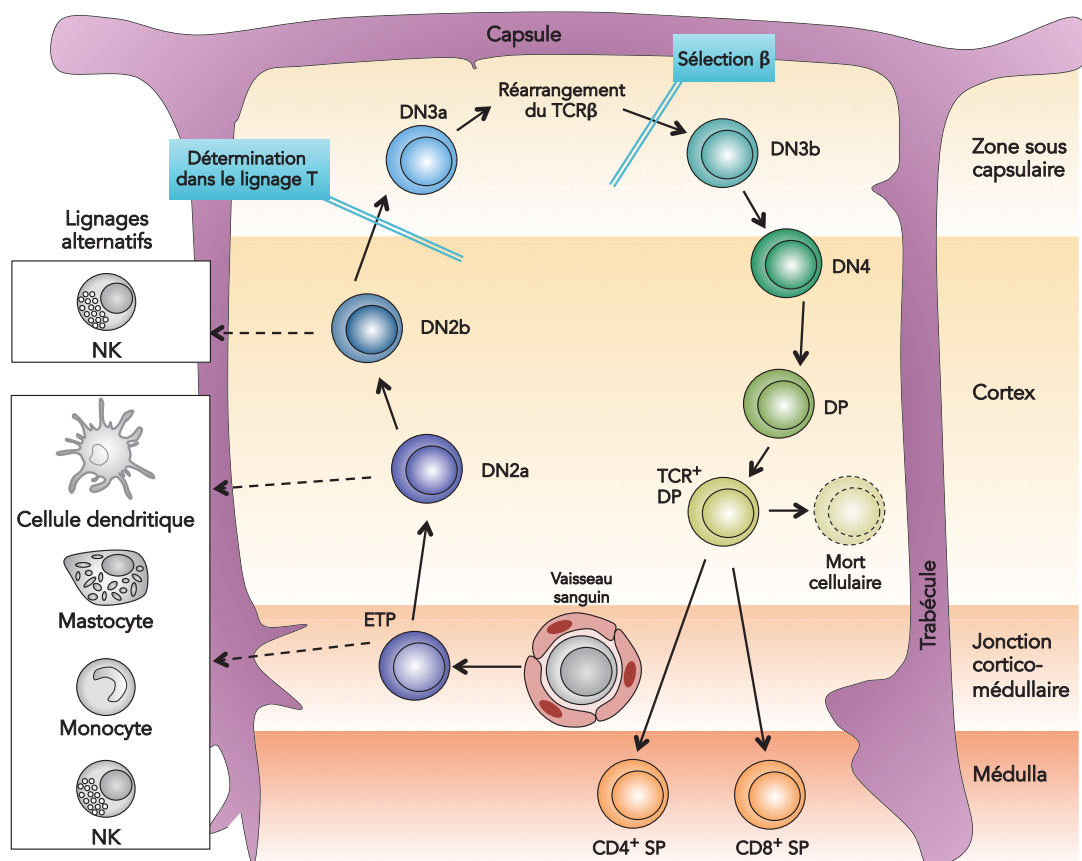


Figure 2 : Développement intra-thymique et détermination des LT

Représentation du développement intrathymique des LT, du stade de progéniteur multipotent au stade CD4 ou CD8 SP. Jusqu'au stade DN2b, les thymocytes sont encore multipotents et peuvent s'engager dans des lignages alternatifs. L'expression des facteurs de transcription BCL11, TCF-1 et GATA-3 initie l'engagement des thymocytes dans la lignée lymphoïde. L'expression des recombinaisons RAG-1 et RAG-2 permet le réarrangement des gènes codant pour la chaîne β du TCR. La signalisation du complexe pré-TCR permet la « sélection β » des thymocytes et marque leur engagement définitif dans le lignage T. Au stade DP, les thymocytes réarrangent la chaîne α du TCR et subissent une étape de sélection positive visant à induire la survie des cellules exprimant un récepteur à l'antigène fonctionnel et restreint pour le CMH du soi. Les précurseurs T utiles pour l'organisme s'engagent alors, possiblement de façon aléatoire, dans le lignage T CD4 ou T CD8 et migrent dans la médulla. Au contact des cellules stromales, ils subissent enfin un processus de sélection négative visant à éliminer les cellules auto-spécifiques avant de finir leur maturation et de migrer dans les organes lymphoïdes périphériques.

II. Homéostasie des lymphocytes T CD4 en périphérie

A. Maturation, trafic et survie des lymphocytes T CD4 naïfs

Les lymphocytes T CD4 qui ont récemment émigré du thymus, ou RTE pour *Recent Thymic Emigrant*, sont qualifiés de naïfs (LT_N) comme toutes les cellules T qui n'ont jamais été activées par une cellule présentatrice d'antigènes mature exprimant à sa surface le ligand de leur TCR. Cependant, si elles sont phénotypiquement proches des LT_N qui patrouillent dans les tissus périphériques depuis plusieurs jours ou semaines, ces cellules ne sont pas fonctionnellement identiques. Afin de les étudier, Pamela Fink a utilisé les souris transgéniques Rag2-GFP initialement générées par le groupe de Michel Nussenzweig (22). Ces souris expriment la molécule fluorescente GFP sous le contrôle du promoteur du gène *Rag2* qui est, comme décrit ci-dessus, activé de façon transitoire dans les thymocytes en cours de développement. Dans ce modèle, la demi-vie de la GFP est d'environ 56h ce qui permet de distinguer les RTE (GFP+) des LT_N matures qui ne l'expriment plus (23, 24). L'utilisation de ces animaux transgéniques a permis de montrer que les RTE ($CD4^+CD62-L^+GFP^+$) sont fonctionnellement distincts des LT_N matures ($CD4^+CD62-L^+GFP^-$). Dans des systèmes de culture *in vitro* permettant de contrôler très finement l'activation et la différenciation des lymphocytes T, les RTE présentent en effet une réponse quantitativement et qualitativement biaisée par rapport aux cellules T matures (25). Ils sont en effet incapables de proliférer et de se différencier en cellules Th1 ou Th17 mais sont par contre plus efficaces que leur contrepartie mature pour acquérir un phénotype Th2 en présence d'IL-4. Dans les organes lymphoïdes périphériques, les RTE subissent donc un processus de maturation qui permet de moduler leur réponse aux signaux environnementaux. Les mécanismes moléculaires qui contrôlent cette maturation terminale et qui pourrait être à l'origine du biais Th2 restent cependant à identifier.

Les LT_N circulent en périphérie entre les organes lymphoïdes secondaires (OL-II), tels que les nœuds lymphatiques ou la rate, via le sang et la lymphe. Dans la circulation sanguine, les LT_N expriment fortement le récepteur aux chimiokines CCR7 qui leur permet de remonter le gradient de chimiokines CCL19 et CCL21 vers les OL-II. Ils y entrent ensuite via des capillaires afférents en ce qui concerne la rate, ou par des vaisseaux sanguins spécialisés, appelés HEV (*High Endothéliale Venules*), en ce qui concerne les nœuds lymphatiques. L'entrée s'effectue par un processus, la diapédèse, que l'on peut décomposer en quatre étapes et qui repose sur la signalisation de récepteurs aux chimiokines et aux intégrines et sur l'interaction entre les molécules d'adhérence présentes à la surface des LT et des cellules endothéliales. *In fine* la

diapédèse permet la transmigration des LT entre les cellules endothéliales. Les OL-II sont structurés en zones T et B qui accumulent respectivement les LT et les LB. Pour accéder aux zones T, les LT se déplacent le long de cellules fibroblastiques réticulaires (FRC) qui sont à l'origine de la production de CCL19 et de CCL21. Au fur et à mesure de leur progression dans l'OL-II, CCR7 est désensibilisé et son expression est diminuée. Cette désensibilisation s'accompagne de l'augmentation de l'expression du récepteur S1PR1 permettant aux LT de suivre le gradient du sphingolipide S1P présent dans la lymphe pour sortir de l'OL-II par les vaisseaux lymphatiques efférents (**Figure 3**). Les LT_N peuvent ainsi patrouiller dans les OL-II pendant plusieurs semaines à la recherche de l'antigène dont ils sont spécifiques.

En l'absence d'activation, les LT_N qui patrouillent dans les organes lymphoïdes secondaires doivent intégrer des signaux de survie afin de pouvoir persister en périphérie pendant plusieurs semaines. Ils leur sont notamment fournis par les DC présentes dans les zones T. Les LT_N établissent avec ces cellules de nombreux contacts de courte durée au travers d'interactions « non-agonistes » entre leur TCR et les complexes CMH-II-peptide qu'elles présentent à leur surface (26). Il a été montré que la faible activation de la signalisation TCR qui en résulte, appelée tonus TCR, est essentielle à la survie des LT_N. En effet, dans un modèle murin dans lequel l'expression des molécules du CMH-II est restreinte aux cellules du stroma thymique, les LT CD4 se développent normalement mais ne peuvent survivre en périphérie. La survie des LT CD4 et leur colonisation en périphérie sont en revanche maintenues lorsque l'hôte exprime le CMH-II à la surface des DC (27). Les LT_N reçoivent également des signaux de survie des FRC qui participent ainsi au maintien de la population des LT CD4 périphérique via la production d'IL-7 (28). L'IL-7 est également produite par les cellules endothéliales lymphoïdes au niveau du sinus sous-capsulaire (29). Cette cytokine signale via l'IL-7R qui est composé d'une chaîne α spécifique à la reconnaissance de l'IL-7 (IL-7R α ou CD127) et de la chaîne commune γ , aussi impliquée dans les événements de signalisation induit par l'IL-2, l'IL-4, l'IL-9, l'IL-15 et l'IL-21 (30). La liaison de la cytokine à l'IL-7R active deux voies majeures. Les kinases JAK1 et JAK2, associées au récepteur, s'activent par *trans*-phosphorylation puis activent en aval le facteur de transcription STAT-5. STAT-5 induit alors l'expression de la molécule anti-apoptotique BCL-2 (B Cell Lymphoma 2) et à l'inverse réprime celle de la molécule pro-apoptotique BIM (BCL-2 Interacting Mediator of cell-death) (31). D'autre part, la signalisation de l'IL-7R permet d'activer la cascade PI3K/AKT/mTOR (Mammalian Target Of Rapamycine) qui induit notamment l'expression du transporteur de glucose GLUT1 et participe ainsi à la fourniture en énergie de la cellule (32, 33).

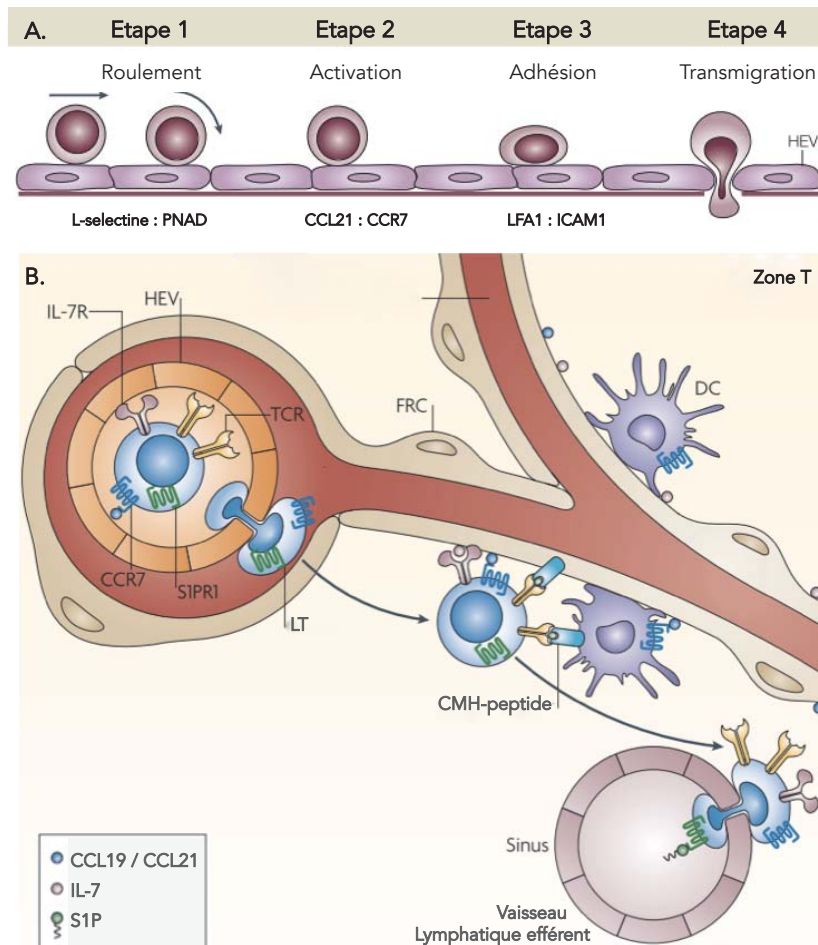


Figure 3 : Circulation des LT CD4 dans les nœuds lymphatique

A) Les LT_N entrent dans les NL par les HEV selon un processus de diapédèse qui se décompose en quatre étapes. 1) Dans la circulation sanguine les LT_N expriment la molécule d'adhésion L-selectine qui reconnaît PNAD (Peripheral Nod Adressin) à la surface des cellules endothéliales ce qui permet aux LT_N de s'attacher et de rouler le long de l'endothélium vasculaire. 2) Les chimiokines CCL21 et CCL19 présentes dans les HEV signalent via CCR7 dans les LT_N ce qui induit l'expression de l'intégrine LFA-1 (Leucocyte Function-associated Antigen 1). 3) L'interaction de LFA-1 avec ICAM-1 (Intracellular Adhesion Molecule), une protéine de la superfamille des immunoglobuline exprimée par les cellules endothéliales, permet l'arrêt ferme des LT sur la paroi vasculaire. 4) Les LT_N remontent le gradient des chimiokines CCL21 et CCL19 produites par les cellules fibroblastiques réticulaires (FRC) et migrent entre deux cellules endothéliales jusque dans la zone T des NL. B) Dans les zones T des OL-II, les LT_N se déplacent le long des FRC et captent l'IL-7 qu'elles produisent. La signalisation de l'IL-7 permet l'induction de molécules anti-apoptotiques telles que BCL-2 dans les LT_N . Les LT_N interagissent également avec des DC ce qui permet de maintenir un tonus TCR nécessaire à leur survie. Les LT_N expriment finalement S1PR1 ce qui leur permet de reconnaître le sphingolipide S1P et de sortir par les vaisseaux lymphatiques efférents. Adapté de (34, 35)

B. Mécanismes de tolérance périphérique

Dans le thymus, la sélection négative est un mécanisme de tolérance qui permet d'éliminer physiquement ou fonctionnellement les thymocytes autospécifiques. Cependant, cette étape de sélection n'est pas sans faille. Il existe en effet des trous dans le répertoire d'antigènes du soi présentés aux cellules T en cours de développement par les mTEC et les DC de la médulla thymique. Des lymphocytes T autospécifiques peuvent donc échapper au processus de sélection et migrer en périphérie. Le risque est alors que ces lymphocytes s'activent et exercent leur fonction immunitaire contre des tissus sains de l'organisme. Pour contrôler cette autoréactivité qui pourrait être à l'origine de pathologies auto-immunes, différents mécanismes de tolérance périphérique sont mis en place dans les organes lymphoïdes secondaires. Ils peuvent être subdivisés en mécanismes « actifs » et « passifs » et ils peuvent aboutir à la mort programmée par apoptose des lymphocytes T effecteurs autospécifiques, à leur anergie, ou à leur différenciation en cellules T régulatrices.

L'anergie est un état de non-réponse prolongé des cellules à leur environnement qui peut être induit suite à la stimulation des cellules T par des cellules présentatrices d'antigènes non matures (36). En condition non-inflammatoire, les cellules dendritiques présentes dans les tissus de l'organisme captent des antigènes du soi sous forme principalement de corps apoptotiques qu'elles internalisent par phagocytose. Elles migrent ensuite dans les zones T des organes lymphoïdes secondaires pour les présenter sous forme de peptides complexés aux molécules du CMH. En l'absence de danger, les cellules dendritiques évoluent dans un environnement non inflammatoire. Elles ne reçoivent donc pas les signaux environnementaux nécessaires à leur maturation et sont en conséquence dans un état de semi-maturité qui ne leur permet pas d'activer efficacement les lymphocytes T. En effet, si elles peuvent leur fournir un signal antigénique, les cellules dendritiques sont par contre incapables de leur délivrer les informations nécessaires à leur reprogrammation métabolique, à leur survie ou à leur différenciation. En l'absence de signal de « co-stimulation », le programme d'expression génique coordonné par le facteur de transcription canonique NFAT, mobilisé en aval du TCR, verrouille le lymphocyte T dans un état d'anergie notamment caractérisé par l'expression des facteurs de transcription EGR2 et EGR3 (Early Growth Response) et des E3 Ubiquitine Ligases CBL-B, GRAIL et ITCH. Ces dernières inhibent la signalisation du TCR à différents niveaux et préviennent la réorganisation du cytosquelette d'actine nécessaire à la prolifération cellulaire (37). Les lymphocytes T auto-spécifiques « activés » par les DC semi-matures sont ainsi « anergisés » puisqu'ils perdent leur capacité à s'activer, à soutenir leur prolifération et à se différencier. Pour mettre en évidence ce mécanisme de tolérance passive, le groupe du Dr.

Nussenzweig a développé un modèle permettant de délivrer de façon spécifique un antigène, le lysozyme d'œuf de poule (HEL) par exemple, aux cellules dendritiques conventionnelles de type 1 (cDC1, voir ci-dessous). Ils l'ont ainsi couplé à un anticorps dirigé contre DEC-205, un récepteur endocytaire exprimé à haut niveau par cette sous-population de DC (38). Ce système offre la possibilité de délivrer et de faire présenter un antigène aux DC sans modifier leur distribution, leur maturation ou leur environnement. Après injection du tandem à des souris de laboratoire et transfert adoptif de lymphocytes T CD4 3A9, qui expriment un TCR spécifique d'un peptide dérivé du HEL présenté par la molécule I-A^k, les auteurs ont observé que les cellules T transgéniques proliféraient faiblement mais ne se différenciaient pas en cellules T effectrices puisqu'elles ne produisaient ni IFN γ ni IL-4. De façon remarquable, la vaste majorité de ces cellules meurt après quelques semaines et les cellules résiduelles sont incapables de répondre à un nouveau challenge avec l'antigène en présence d'un adjuvant fort. La présentation d'antigènes par les cDC1 dans un contexte non inflammatoire peut donc aboutir à l'anergie des cellules T. Le groupe du Dr. Nussenzweig a donc mis en évidence que les DC peuvent avoir un rôle tolérogène. Il a depuis été montré que la présentation d'antigène du soi par ces DC semi-matures participe à la tolérance périphérique en induisant l'anergie des LT auto-spécifique mais aussi en induisant leur mort ou leur différenciation en lymphocytes T régulateurs doués d'activité immunosuppressive (39). Il semble que le devenir des LT stimulés dans ce contexte tolérogène dépende de la population de DC impliquée et des signaux moléculaires présents dans l'environnement, notamment ceux comme le TGF- β ou l'acide rétinoïque qui instruisent la différenciation en Treg (40, 41).

En complément de ces mécanismes de tolérance passive, des mécanismes qualifiés d'actifs, car reposant sur l'action de cellules T régulatrices Foxp3⁺, sont nécessaires à la régulation de l'homéostasie du compartiment lymphocytaire T périphérique et au contrôle des clones autospécifiques. Le rôle critique de ces cellules est illustré par le fait que des souris dont le gène *Foxp3* n'est pas fonctionnel, et qui n'ont donc pas cette population régulatrice, meurent rapidement après la naissance d'un syndrome auto-immun multi-organes (42). Chez l'Homme, des mutations perte de fonction du gène *Foxp3* entraînent aussi l'apparition d'un syndrome auto-immun sévère appelé IPEX (Immunodysregulation Polyendocrinopathy Enteropathy X-linked) (43). Afin de contrôler le répertoire T, notamment autospécifique, les lymphocytes T régulateurs utilisent un large répertoire de mécanismes effecteurs. Ils sécrètent par exemple des cytokines immunosuppressives, telles que le TGF- β ou l'IL-10, qui inhiberont l'activation et la différenciation des LT effecteurs de façon directe ou indirecte, en réprimant par exemple la maturation des DC (44). La plupart des lymphocytes T régulateurs exprimant de façon constitutive CD25, et donc un récepteur de haute affinité pour l'IL-2, ces cellules régulent aussi l'activité des cellules effectrices en purgeant cette cytokine lymphoproliférative

de leur environnement (44). Parmi les mécanismes immunosuppresseurs les mieux caractérisés, il a aussi été montré que les cellules régulatrices peuvent induire l'expression de l'indoléamine 2, 3-dioxygénase (IDO) par les DC via l'interaction de leur molécule de surface CTLA-4 avec les molécules CD80/86 exprimées à la surface des DC (45). La surexpression d'IDO induit le catabolisme du tryptophane ce qui inhibe la prolifération des cellules T et entraîne leur mort. L'un des catabolites de cet acide aminé essentiel, la kynurénine, est en effet pro-apoptotique.

III. Réponses des lymphocytes T CD4 en périphérie

A. Les cellules dendritiques : un lien entre immunité innée et adaptative

Dans les tissus périphériques, lymphoïdes ou non, les DC jouent le rôle de sentinelles. Elles ont en effet pour principale vocation de détecter toute perturbation de l'homéostasie tissulaire par des dangers endogènes ou exogènes, d'intégrer les signaux innés produits par les tissus et les autres cellules de l'immunité, et de les traduire dans un langage interprétable par les cellules T. Pour détecter le danger, les DC expriment des récepteurs innés appelés PRR (Pattern Recognition Receptor). Contrairement au TCR ou au BCR, ces récepteurs sont encodés par des gènes intacts de la lignée germinale. Ils détectent une grande diversité de motifs, qualifiés de PAMP (Pathogen-Associated Molecular Pattern), conservés au cours de l'évolution et partagés par différents types d'agents pathogènes (46, 47). En effet, les PRR reconnaissent des motifs communs à divers dangers exogènes tels que des acides nucléiques, des sucres ou des lipopeptides. Comme initialement postulé, les PRRs peuvent aussi reconnaître des DAMP (Damaged-Associated Molecular Pattern) (48). Il s'agit pour l'essentiel de molécules préformées et séquestrées à l'intérieur des cellules de l'organisme qui sont relarguées dans les tissus lorsqu'elles meurent par nécrose ou en situation de stress (49). Les protéines de choc thermique (HSP), la protéine nucléaire HMGB1 et l'acide urique font partie des DAMPs les mieux caractérisés. Il existe quatre grandes familles de PRR : les récepteurs Toll Like (TLR), les lectines de type C (CLR), les récepteurs NOD-like (NLR) et les récepteurs RIG-I-like (RLR). Toutes ces familles comprennent plusieurs membres qui reconnaissent des PAMP ou DAMP différents. Les différents PRR signalent aussi par des voies au moins en partie distinctes et induiront donc des programmes d'expression génique différents. Si la plupart des TLR active les voies MAPK ou NFκB, la capacité à mobiliser la voie IRF3 qui *in fine* aboutit à la production d'IFN de type I est par exemple restreinte aux récepteurs couplés à TRIF, tels que TLR3 et

TLR4. Pour renforcer l'efficacité de ce système de détection, les PRRs ont de plus des localisations subcellulaires différentes. La plupart des TLR et les CLR sont ainsi présents à la membrane plasmique mais on trouve aussi certains de ces récepteurs reconnaissant des acides nucléiques tels que TLR3, TLR7 ou TLR9, dans les endosomes. Enfin, les NLR et les RLR MDA5 et RIG-1 sont localisés dans le cytoplasme.

Chaque agent pathogène va donc être détecté par un panel restreint de PRR qui va donc agir comme un code barre. Ils vont de façon concomitante ou séquentielle, déclencher un programme d'expression génique qui aboutira de façon quasi-systématique à la sur expression des molécules impliquées dans le préparation de l'antigène, la présentation du peptide antigénique et dans la communication avec les cellules T (e.g. CMH I et II, molécules de co-stimulation CD80/86, CD40). Chaque pathogène déclenchera aussi la production d'un set restreint de médiateurs solubles, notamment de cytokines inflammatoires, qui joueront un rôle clef dans la polarisation des réponses T CD4.

L'intégration des signaux de danger par les DC permet donc d'induire un programme de maturation qui augmente de façon considérable leur antigénicité et leur immunogénicité (50). Cette maturation est un processus hiérarchisé dans le temps et l'espace, depuis les tissus périphériques jusqu'aux organes lymphoïdes drainant où les DC vont interagir avec les cellules T afin de déclencher la phase adaptative de la réponse immunitaire.

B. L'activation et la différenciation des LT CD4 nécessitent 3 signaux

Afin de protéger l'organisme contre de multiples dangers, les lymphocytes T CD4 naïfs « scannent » les cellules dendritiques dans les organes lymphoïdes secondaires « à la recherche » des complexes CMH-II-peptide dont ils sont spécifiques. Pour s'activer et se différencier, les lymphocytes T CD4 naïfs ont besoin d'intégrer trois types de signaux : un signal antigénique qu'ils captent via leur TCR, un signal de co-stimulation impliquant le plus souvent la molécule CD28, et un signal polarisant pouvant être exprimé et sécrété par les cellules dendritiques, des cellules tierces, ou les lymphocytes eux-mêmes. L'activation des lymphocytes T naïfs, ou *priming*, se déroule dans les zones T des organes lymphoïdes secondaires et peut se décomposer dans le temps en trois phases. Tout d'abord, le lymphocyte T scanne les complexes CMH-II-peptide T et établit pour ce faire des interactions de courte durée et de faible affinité avec les cellules dendritiques. La durée des interactions augmente progressivement de façon concomitante à l'augmentation de l'expression des

marqueurs d'activation comme CD25, CD69 et CD44 et à un ralentissement de la motilité des lymphocytes T CD4 naïfs. Lors de la seconde phase, ils interagissent durablement avec des cellules dendritiques matures à la surface desquelles ils ont détecté le ligand de leur TCR. C'est lors de cette interaction soutenue que les lymphocytes T CD4 naïfs s'activent et se différencient en cellules effectrices. Une des conséquences de l'activation est l'expansion clonale des lymphocytes T. Cette dernière est concomitante à la troisième phase du modèle cinétique. Lors de cette prolifération importante les lymphocytes T regagnent en vélocité et les interactions avec les cellules dendritiques deviennent de courtes durées jusqu'au détachement complet des lymphocytes et des cellules dendritiques (51).

L'activation et la différenciation des lymphocytes T nécessitent l'intégration de différents signaux par les récepteurs membranaires au niveau d'une structure anatomique particulière, appelée synapse immunologique, qui s'organise à l'interface entre la DC et le LT lorsque ces deux populations cellulaires interagissent de façon stable et durable (**Figure 4**). Leur signalisation se traduit par la mobilisation et la coopération de nombreux acteurs moléculaires intracellulaires regroupés au sein d'un complexe supramoléculaire d'activation (SMAC). Le centre de ce complexe (cSMAC) regroupe les récepteurs et leurs molécules de signalisation nécessaires à la survie, l'activation, la prolifération et la différenciation des LT. En périphérie du complexe (pSMAC), la liaison des molécules LFA-1 à la surface de LT avec ICAM-1 à la surface des DC permet d'établir des jonctions adhérentes stables entre les deux cellules. L'exclusion d'un récepteur au-delà de la synapse immunologique (dSMAC) empêche son interaction avec son ligand et sa signalisation. C'est par exemple le cas du corécepteur CD28 ou du récepteur à l'IFN- γ qui peuvent être exclus au niveau du dSMAC par la molécule inhibitrice CTLA-4 (Cytotoxique T Lymphocytes Antigen 4) et par la signalisation de l'IL-4 respectivement (26, 52, 53).

Au sein de la synapse immunologique, les lymphocytes T CD4 naïfs vont donc recevoir trois types de signaux leur permettant de s'activer et de se différencier dans le sous-type de lymphocytes T effecteurs adapté à l'élimination du danger détecté par les cellules dendritiques. Le signal TCR et les signaux de costimulation permettent notamment l'activation, la survie et la prolifération cellulaires. Les signaux polarisant instruisent quant à eux le lignage de lymphocytes effecteurs.

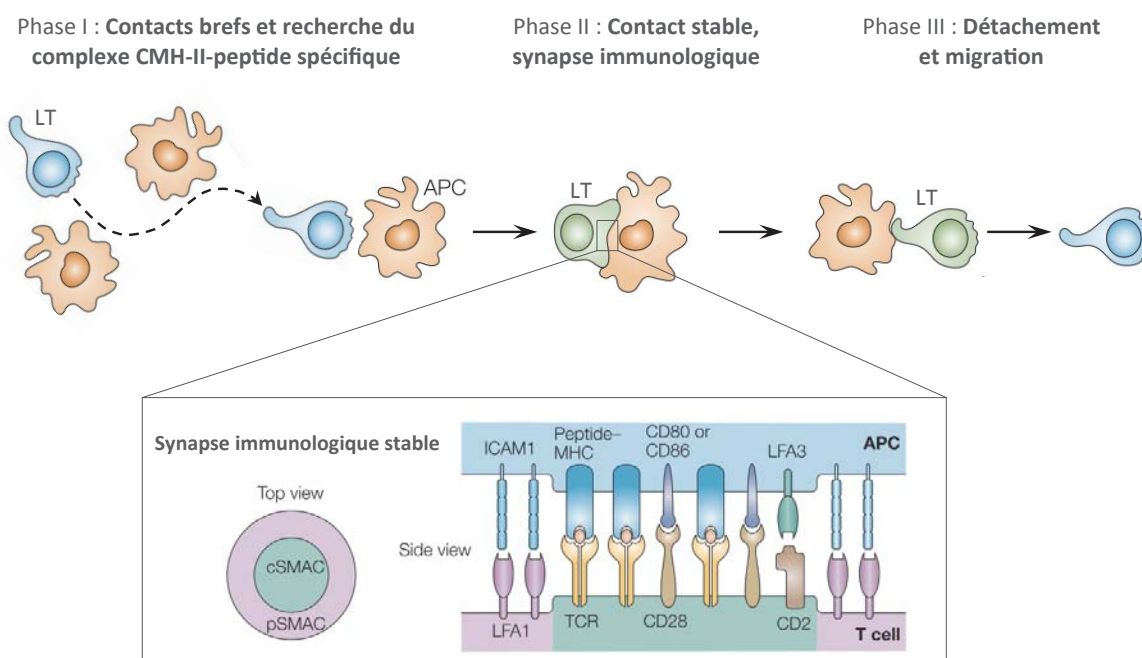


Figure 4 : Modèle d'interaction des lymphocytes T avec les cellules dendritiques et synapse immunologique

Les lymphocytes « scannent » les cellules dendritiques en établissant des contacts brefs à la recherche des complexes CMH-II-peptides dont ils sont spécifiques. Les deux cellules établissent ensuite un contact stable via la formation d'une synapse immunologique. Elle est organisée en complexe supramoléculaire d'activation qui regroupe les molécules du complexes TCR et des molécule d'adhérence entre autres en son centre et qui stabilisée dans se périphérie par d'autre molécules d'adhérence. Enfin les deux cellules se détachent, les lymphocytes T prolifèrent et migrent vers leurs tissus cibles. Les LT illustrés en bleu symbolisent les lymphocytes « mobiles » par opposition à ceux illustré en vert qui sont peu mobiles et signalisent intensément. Adapté de (26).

1. Le signal TCR

Comme expliqué précédemment, le récepteur à l'antigène des lymphocytes T est généré dans les thymocytes en cours de développement par recombinaison somatique aléatoire des segments de gène des *loci Tcra* et *Tcrb*. Il est donc exprimé de façon clonale ce qui permet aux cellules T CD4, à l'échelle de leur large répertoire, de reconnaître l'ensemble des dangers qui menacent l'intégrité de l'organisme. Lors de l'intrusion d'un agent pathogène par exemple, seuls les lymphocytes T spécifiques de ce microorganisme doivent s'activer pour éviter la mise en place de réponses immunitaires délétères ou non protectrices. Le verrou qui garantit la spécificité antigénique de la réponse est représenté par l'interaction entre le TCR et le

complexe CMH II-peptide dont il est spécifique. Seuls les lymphocytes T qui reconnaissent le ligand de leur TCR avec une affinité suffisante sont en effet capables d'interagir longuement avec la cellule présentatrice d'antigène, et donc d'intégrer qualitativement et quantitativement les signaux nécessaires à leur activation, prolifération et différenciation. Les mécanismes centraux et périphériques d'induction de tolérance, passifs comme actifs, garantissent quant à eux que les cellules T exprimant un TCR spécifique d'antigène du soi ne puissent s'activer dans un contexte inflammatoire, même si leur ligand est exposé à la surface d'une DC mature.

Les régions intracytoplasmiques des chaînes alpha et bêta du TCR étant extrêmement courtes, le récepteur à l'antigène doit se complexer avec des modules de signalisation pour convertir le signal environnemental détecté en événements de signalisation intracellulaire. Il s'agit de trois dimères de protéines transmembranaires CD3 (CD3_{TC}, CD3_{BC}, and CD3_{EC}) qui, dans les secondes suivant l'interaction du TCR avec le complexe CMH-II-peptide, vont être phosphorylées au niveau de leurs motifs ITAM (Immunoreceptor tyrosine-based activation motif) par la src kinase p56^{Lck} arrimée au domaine intracellulaire du co-récepteur CD4. Les motifs ITAM phosphorylés vont alors permettre le recrutement de la kinase ZAP70 par son tandem de motifs SH2. Après avoir été elle-même activée par p56^{Lck}, ZAP70 phosphoryle à son tour SLP76 (*SH2-domain-containing leukocyte protein of 76 kDa*) et LAT (*Linker for the activation of T cells*). Ces deux protéines servent alors d'échafaudage, de points d'ancrage à de nombreux intermédiaires de signalisation qui allumeront les différentes voies, souvent interconnectées, nécessaires à l'activation des lymphocytes T (54) (**Figure 5**).

Parmi les intermédiaires de signalisation la phospholipase Cγ (PLCγ) joue un rôle crucial. Elle est à la base de l'activation de trois facteurs de transcription canoniques : AP1 (Activator Protein 1), NFκB (nuclear factor-κB) et NFAT (nuclear factor of activated T cells). En effet elle permet de dégrader le phosphatidylinositol 4,5-bisphosphate en inositol trisphosphate (InsP3) et en diacylglycerol (DAG). Le premier, en se liant sur son récepteur (IP3R1, pour Inositol 1,4,5-trisphosphate receptor type 1) au niveau de la mitochondrie induit notamment un influx calcique dans la cellule ce qui active la calcineurine. Cette phosphatase, en déphosphorylant NFAT, permet l'activation du facteur de transcription qui est importé dans le noyau. Le DAG permet quant à lui l'activation de la PKCθ (Protéine Kinase Cθ) entre autres. Cette dernière phosphoryle la protéine IκB (inhibitor kappa-B) impliquée dans la séquestration dans le cytoplasme de NFκB. Une fois phosphorylée IκB se dissocie de NFκB ce qui permet la translocation nucléaire du facteur de transcription. Le DAG permet également l'activation des kinases ERK (extracellular signal-regulated kinase) et MAPK (Mitogen activated protein kinase) qui phosphorylent notamment la sous-unité c-Jun du facteur de transcription canonique AP-1 et permet ainsi son activation.

(54)

(54)

(54)

2. Le signal de co-stimulation

L'interaction de la protéine de surface CD28 avec les molécules de co-stimulation CD80/CD86 (B7.1 et B7.2), uprégulées à la surface des DC après maturation, va fournir le deuxième signal nécessaire à l'activation et la différenciation des LT_N. La principale voie de signalisation activée en aval de CD28 dépend de la Phosphoinositide 3-Kinase (PI3K). Dans les secondes suivant l'engagement de CD28, la phosphorylation de la tyrosine du motif YMMN par les src kinases p56^{lck} et p59^{fyn} permet son recrutement via le domaine SH2 de sa sous-unité régulatrice p85. La PI3K ainsi amarrée à la région intracytoplasmique de CD28 va phosphoryler le Phosphatidyl-Inositol bi-Phosphate (PIP2) et ainsi produire du PIP3 qui va recruter, via leur domaine Pleckstrin Homology (PH), diverses protéines de signalisation dont PDK1 et PKB/AKT. PDK1 ainsi dockée à la membrane de la cellule T va phosphoryler et activer AKT mais aussi S6K et la PKC θ . AKT va alors à son tour, directement ou indirectement, allumer de nouvelles voies de signalisation ou booster celles mobilisées par le signal TCR.

Parmi les nombreuses fonctions cellulaires régulées par CD28, il a été montré que cette molécule jouait un rôle clef dans la survie des cellules T. Cette fonction dépend notamment de la capacité d'AKT à inhiber la protéine pro-apoptotique BAD (*Bcl-2-associated death promoter*) et à induire, conjointement avec la PLC γ , la surexpression de la molécule anti-apoptotique BCL-X_L via la mobilisation d'NF κ B (55). De façon conjointe avec AP-1 et NFAT, ce facteur de transcription induit aussi la production d'IL-2 par les cellules T et l'expression de CD25, la chaîne α du récepteur de haute affinité à cette cytokine lymphoproliférative. Cette cascade moléculaire conduit à l'activation de la voie de signalisation de l'IL-2 au sein des LT de manière autocrine et paracrine ce qui booste leur prolifération (56, 57). Il faut cependant noter que l'impact de CD28 sur l'entrée et la progression des cellules dans le cycle cellulaire n'est que partiellement dépendant de l'IL-2. En effet Appleman et collaborateurs ont montré que la signalisation de CD28 potentialise aussi directement l'expression des cyclines D2 et E (58).

La voie CD28/PI3K/AKT est aussi essentielle à la reprogrammation métabolique des cellules T après activation. Il s'agit d'un rôle majeur car les besoins énergétiques des cellules vont considérablement augmenter puisqu'elles devront rapidement assumer la synthèse des composés cellulaires nécessaires à leur expansion clonale et la sécrétion de quantités importantes de médiateurs solubles. Comme discuté précédemment, les LT_N sont des cellules quiescentes qui n'ont donc pas besoin d'activer les voies anaboliques nécessaires à la synthèse *de novo* d'ADN, de protéines ou de lipides. Leurs faibles besoins énergétiques peuvent être couverts par un catabolisme des nutriments reposant sur la phosphorylation oxydative. Après

activation, les cellules subissent une reprogrammation métabolique pour assumer leurs besoins croissants en énergie (32). Dans ce contexte, la PI3K a un rôle clef puisqu'elle permet notamment d'activer la voie mTOR en inhibant un inhibiteur de Rheb. Frauwirth et collaborateurs ont aussi montré que l'activation de cette cascade moléculaire permettait d'augmenter l'expression du récepteur au glucose GLUT1 et ainsi d'augmenter l'apport énergétique de la cellule (59).

Il est important de noter que la voie de co-stimulation reposant sur l'interaction de CD80/86 avec CD28 n'est pas la seule à avoir été identifiée dans la littérature. Il a en effet été montré que l'axe C70/CD27 peut aussi avoir un rôle clef dans le priming des lymphocytes T CD4 naïfs. CD27 fait partie de la famille des récepteurs au TNF et est constitutivement exprimé à la surface des LT_N, CD4 comme CD8. Son ligand, CD70, est quant à lui surexprimé à la surface des DC matures. En utilisant un modèle de souris déficientes pour CD27 et/ou CD70, le groupe de Jannie Borst a notamment montré que CD27 était nécessaire à la survie et l'expansion des cellules T CD4 et CD8, et qu'en son absence la réponse immunitaire de ces deux populations cellulaires en réponse à une infection par le virus influenza était fortement diminuée (60, 61). De façon remarquable, ces travaux ont de plus montré que CD27 et CD28 assument des fonctions au moins en partie non redondantes. L'impact de CD27 sur l'expansion des cellules T est notamment indépendant de l'effet de CD28 sur le cycle cellulaire et de la voie IL-2 (60).

3. Le signal polarisant

Comme discuté précédemment, les lymphocytes T CD4 naïfs sont des cellules extrêmement plastiques, capables d'adapter leur phénotype et leurs fonctions au type de danger détecté par les cellules de l'immunité innée et des tissus. Il a par exemple été montré que les LT_N peuvent se différencier en cellules Th1 qui, via une production massive d'IFN- γ , jouent un rôle clef dans la coordination des réponses immunitaires dirigées contre des tumeurs ou contre des pathogènes intracellulaires. Le potentiel de différenciation des lymphocytes T naïfs est vaste et leur permet d'acquérir de multiples phénotypes. Les principales sous-populations de lymphocytes T helpers, ainsi que leurs répertoires de facteurs de transcription, de cytokines et de récepteurs aux chimiokines, seront présentés dans la section suivante.

Les LT_N peuvent donc acquérir un large panel de phénotypes afin de protéger au mieux l'organisme contre l'ensemble des dangers qui le menacent, qu'ils soient endogènes ou exogènes. Mais comment le lymphocyte T CD4 connaît-il la voie à emprunter compte tenu du

fait qu'il n'a pas, à quelques rares exceptions près, d'interaction directe avec l'agent pathogène. Même s'ils peuvent influencer son choix, les signaux 1 et 2, communs à tous les lignages, ont un impact limité sur son engagement et sa détermination vers un phénotype donné. Le choix de lignage est en fait dicté à la cellule T par un troisième signal, dont l'origine, la qualité et la quantité dépendent essentiellement du type de danger détecté.

Même s'ils ne sont pas spécifiques d'un pathogène donné, les PRRs déclenchent des réponses immunitaires distinctes selon le type de micro-organisme rencontré. Comme discuté ci-dessus, cette plasticité est liée au fait qu'il existe de nombreuses classes de récepteurs innés, reconnaissant des motifs conservés différents et activant des voies de signalisation distinctes. De plus, en fonction de la cellule qui l'exprime, un même PPR pourra déclencher des programmes d'expression génique partiellement différents. Enfin, parce qu'ils expriment de multiples PAMPs, ou parce qu'ils changent de localisation subcellulaire ou de forme au cours de leur cycle de vie, la reconnaissance des pathogènes par les cellules de l'organisme implique souvent différents types de PRRs. Chaque type de danger est donc détecté par un set restreint de PRRs qui, en réponse à leur engagement, induisent l'expression d'une signature moléculaire précise. Ce sont ces signaux, regroupés sous le terme de « signal 3 », qui dictent le choix de lignage du lymphocyte T CD4.

Le signal 3 repose essentiellement sur des cytokines. Si elles diffèrent d'un lignage à l'autre, le programme qu'elles induisent s'établit selon le même modèle (**Figure 6**). Les facteurs de transcription en aval des récepteurs aux cytokines, qualifiés de « pionniers », vont agir avec les facteurs « canoniques » (e.g. AP-1, NFκB) activés en aval du complexe TCR pour initier la programmation des lymphocytes T. Outre l'activation d'un cœur de programme commun à tous les phénotypes, régulant notamment des grandes fonctions cellulaires comme la prolifération ou la survie, les facteurs pionniers vont aussi participer à la mise en place de « sous-programmes » spécifiques des lignages. La plupart des récepteurs aux cytokines sont couplés à des Janus Kinases (JAKs) qui, lorsque le récepteur lie la cytokine, s'activent par transphosphorylation ce qui permet le recrutement des facteurs de transcription STAT (Signal Transducer and Activator of the Transcription) au niveau du récepteur. La phosphorylation et l'activation des STATs sont assurées par les JAKs, ce qui induit leur translocation nucléaire. Les STAT sont les facteurs de transcription pionniers qui initient les programmes d'expression génique caractéristiques des différents lignages (62, 63). En plus de remodeler le paysage épigénétique des cellules (Cf Chapitre III), chaque STAT induit l'expression d'un set restreint de gènes, dont ceux codant pour le « master régulateur » et pour la (ou les) cytokine(s) caractéristique(s) du lignage. Le master régulateur, et le facteur de transcription mobilisé par la cytokine, induisent enfin l'expression d'une deuxième vague de facteurs de transcription qui

viendront affiner et stabiliser la programmation. Si l'on se réfère au lignage Th1 par exemple, l'IL-12 produite par les cellules dendritiques mobilise ainsi STAT-4 qui induit l'expression du master régulateur T-bet et la production d'IFN- γ . T-bet et STAT-1, suite à son activation par le récepteur IFN- γ , induisent alors l'expression de facteurs de transcription tels que Runx3 ou Hlx qui finissent d'engager le lymphocyte T dans le lignage Th1.

Le modèle décrit ci-dessus, comme toute généralisation, présente des exceptions. La source cellulaire du signal 3 n'est pas systématiquement la cellule dendritique. Les éosinophiles, les cellules NK ou les lymphocytes T CD4 eux-mêmes peuvent par exemple, pour certains lignages et/ou dans certaines situations physiopathologiques, constituer la principale source de cytokine instructrice. Le signal 3 n'est pas non plus systématiquement soluble. Des molécules membranaires, telles que les récepteurs Notch, peuvent aussi dicter le choix de lignage après interaction avec leur ligand, notamment dans le cas de la polarisation Th2 (64). Enfin, le signal 3 peut être modulé par les signaux 1 et 2. Pour exemple, il a récemment été montré qu'un fort signal TCR promouvait la programmation Th1 en induisant la surexpression de la chaîne $\beta 2$ du récepteur à l'IL-12, augmentant ainsi la sensibilité des cellules à cette cytokine pro-Th1 (65). Il a aussi été montré que l'axe de co-stimulation CD70/CD27 pouvait induire la différenciation de cellules T naïve vers un phénotype Th1 de façon indépendante de l'IL-12 (66), mais aussi antagoniser la mise en place du programme Th17 (67).

Les fonctions effectrices des différents lignages s'acquièrent au fur et à mesure du cycle cellulaire. En suivant la prolifération des LT en cours de différenciation Th1 et Th2 par dilution du CFSE (un marqueur intravital), il a été montré que la production des cytokines effectrices du lignage augmentait avec le nombre de cycles (68, 69). De plus le groupe du Dr. Reiner a mis évidence la nécessité de la prolifération pour l'acquisition de fonctions effectrices en bloquant la différenciation des LT par traitement des cellules avec des inhibiteurs du cycle cellulaire (68).

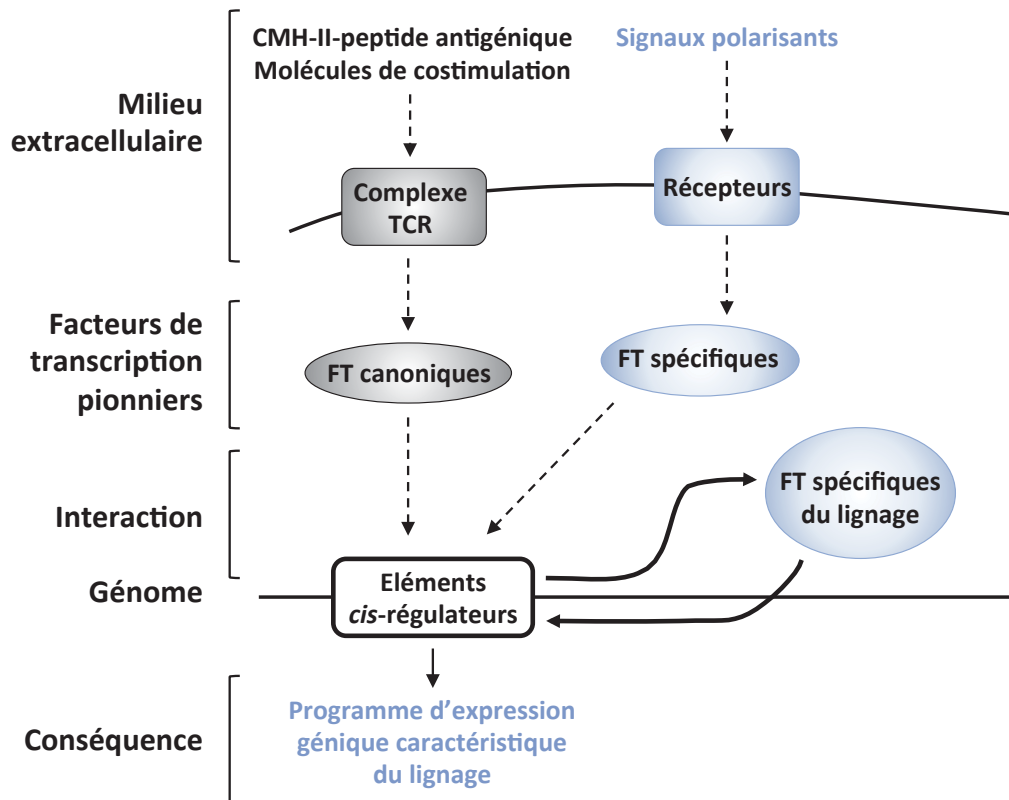


Figure 6 : Modèle de polarisation des lymphocytes T CD4

La signalisation du complexe TCR active des facteurs de transcription « canoniques » tels que AP-1 ou NFkB. Les signaux polarisants présents dans le milieu extracellulaire vont activer des récepteurs dont la signalisation met en jeu des facteurs de transcription spécifiques du signal perçu. Ces facteurs de transcription « pionniers » initient la mise en place du programme d'expression génique caractéristique du lignage dans lequel le lymphocyte T s'est engagé. Ils permettent également l'expression de facteurs de transcription spécifiques du lignage qui amplifient et stabilisent la différenciation.

C. Hétérogénéité et plasticité des lignages T helpers

1. Hétérogénéité des lignages T CD4

Un récapitulatif des signaux instructeurs, des sets de facteur de transcription mis en jeu pour instaurer le programme d'expression génique, et des fonctions effectrices de chaque lignage est illustré en **Figure 7**.

a. Les lymphocytes de type Th1

Le lignage Th1 est décrit pour la première fois par Mosman et Coffman en 1986 (70). Il est instruit par l'IL-12 sécrétée essentiellement par les cDC1 en réponse à l'engagement de leur PRR et après leur instruction par les LT CD4 via l'interaction CD40/CD40L (71, 72). L'IL-12 est un hétérodimère composé de la chaîne IL-12p40 (qui peut aussi s'homodimériser ou interagir avec l'IL-12p19 pour former l'IL-23) et de la chaîne IL-12p35 spécifique à l'IL-12. Le récepteur est également composé de deux chaînes. Le modèle accepté est que la chaîne $\beta 1$ du récepteur (IL-12 $\beta 1$) est exprimée de façon constitutive sur les LT_N et les LT de type Th1. L'expression de l'IL-12 $\beta 2$ est par contre induite par l'IL-2, l'IFN- γ (73, 74) et une forte signalisation TCR (65). L'IL-12R signale par le biais du facteur de transcription STAT-4 qui induit l'expression de la cytokine effectrice du lignage : l'IFN- γ (75). En 1996 deux équipes distinctes ont démontré l'importance de STAT-4 dans le priming Th1 en montrant que des LT CD4 déficients pour ce facteur de transcription étaient incapables de produire de l'IFN- γ lorsqu'ils étaient exposés à l'IL-12 (76, 77). STAT-4 induit également l'expression de Tbet, le « master régulateur » caractéristique du lignage (78). Les deux facteurs de transcription soutiennent la production d'IFN- γ et induisent aussi l'expression de son récepteur (79, 80). L'IFN- γ en plus de son rôle effecteur, agit de manière auto- et paracrine sur les LT de type Th1 en cours de différenciation en activant le facteur de transcription STAT-1. Il peut également être produit par les cellules NK recrutées dans les nœuds lymphatiques pour soutenir la différenciation Th1 (81). La coopération de STAT-1, STAT-4 et Tbet permet l'établissement et la stabilisation du programme Th1 (82–84). Il se caractérise donc par l'expression d'un réseau restreint de facteurs de transcription incluant, outre les molécules décrites ci-dessus, EOMES, IRF1, HLX ou RUNX3. Il comprend aussi l'expression d'un set précis de récepteurs aux chimiokines, notamment CXCR3 et CCR5, qui permettent aux LT de migrer dans les tissus dans lesquels le danger a été détecté (85). Enfin, les cellules Th1 sont caractérisées par la production d'un répertoire limité de cytokines responsables des effets biologiques qui leur sont associés. En plus de l'IFN- γ , il a en effet été reporté que ces cellules sécrètent de l'IL-2, du TNF et du GM-CSF.

Même si l'IL-12 est décrite comme la cytokine « instructrice » du lignage Th1, des travaux du groupe de Ralph Steinman ont montré que cette molécule n'était pas systématiquement nécessaire. Dans un système reposant sur le ciblage d'antigène aux cDC1 à l'aide d'anticorps dirigés contre DEC205, ils ont pu montrer que le priming Th1 induit était indépendant de l'IL12. L'adjuvant utilisé dans ce modèle, le poly I:C, active TLR3 et/ou RIG-I et MDA5 qui à leur tour induisent une forte surexpression de CD70 à la surface des DC (66). En interagissant avec le récepteur CD27 à la surface des LT_N, CD70 induit la production d'IFN- γ par les LT et leur différenciation en cellule de type Th1.

Les lymphocytes de type Th1 sont impliqués dans la défense de l'organisme contre les pathogènes intracellulaires ou encore les cellules tumorales. Ce lignage T helper est en effet le seul à être systématiquement associé à un bon pronostic chez les patients atteints de cancer (86). L'IFN- γ qu'ils produisent favorise le recrutement et l'activation de macrophages en cellules de type M1 produisant de grande quantité de cytokines pro-inflammatoires (e.g. TNF- α et IL-6) et d'espèces réactives de l'oxygène (ROS) (87). L'IFN- γ induit aussi l'expression des molécules de CMH-I à la surface de cellules infectées ou cancéreuses ce qui augmente la probabilité qu'elles soient reconnues par des lymphocytes T CD8 cytotoxiques et lysées (88, 89). L'IFN- γ permet enfin d'amplifier les capacités cytotoxiques des cellules T CD8 et NK.

b. Les lymphocytes de type Th2

A l'instar du lignage Th1, le lignage Th2 est décrit pour la première fois par Mosman et Coffman en 1986 (70). Il est instruit par la cytokine IL-4 (90, 91). La source cellulaire de l'IL-4 *in vivo* reste sujette à investigation et semble différente selon le contexte physiopathologique. Il a été montré qu'elle pouvait être produite par les basophiles suite à une infection par le nématode *Nippostrongylus brasiliensis* (92) ou par des cDC2 dans le cas d'une infection avec *Schistosoma japonica* (93). Il a aussi été proposé que l'IL-4 produite par les mastocytes et les éosinophiles participe à l'initiation de la réponse Th2 (94). L'IL-4 est reconnue par son récepteur (IL-4R) qui est composé d'une chaîne α de haute affinité et de la chaîne commune γ à la surface des lymphocytes T naïfs. A l'inverse du récepteur à l'IL-12, l'IL-4R n'est pas induit par un fort signal TCR et la différenciation des LT_N en LT de type Th2 ne requiert qu'un faible signal TCR. L'IL-4R signale par le biais du facteur de transcription pionnier STAT-6 ; celui-ci joue un rôle primordial dans le développement des LT de type Th2 car des LT *Stat6*^{-/-} ne peuvent produire d'IL-4 (95). De plus l'expression ectopique de STAT-6 dans des LT en cours de différenciation Th1 détourne le processus de différenciation vers le lignage Th2 (96). En collaboration avec STAT-6, le « master régulateur » GATA-3 dont l'expression augmente au

cours de la différenciation induit la production des cytokines effectrices de type Th2 à savoir l'IL-4, l'IL-13 et l'IL-5 (97) ainsi que l'expression des récepteurs aux chimiokines CCR4 et CCR8 (85). Enfin, de la même manière que l'IFN γ au sein des LTh1, l'IL-4 contribue également, via un rétrocontrôle positif, au maintien du lignage Th2 par l'induction des facteurs de transcription tels que GFI1 et c-MAF (98).

Une voie additionnelle de différenciation des lymphocytes T naïfs vers le type Th2, indépendante de la signalisation de l'IL-4R et de STAT-6, a été mise en évidence par le groupe du Dr. Flavell. Elle implique la signalisation de récepteur Notch1/2 exprimé par les LT_N. Celle-ci est induite après interaction avec Jagged1/2, qui sont deux protéines de surface exprimées par les APC après activation par des allergènes ou des parasites. L'interaction ligand/récepteur induit le clivage du domaine intracellulaire de Notch1/2 qui se lie aux cofacteurs MAML1 (Mastermind-Like 1) et RBPJ (Recombination-signal Binding Protein for immunoglobulin kappa J region). Ce complexe tripartite est ciblé sur le promoteur de *Gata3* et sur l'enhancer HS5 de l'*Il4* par RBPJ ce qui induit l'expression de ces deux gènes et initie la différenciation Th2 (64).

Les cytokines effectrices de type Th2 (IL-4, IL-13 et IL-5) exercent des effets pléiotropiques et confèrent aux lymphocytes T de type Th2 un rôle central dans la défense de l'organisme contre les parasites extracellulaires tels que les Helminthes (99). Les récepteurs à l'IL-4 et l'IL-13 possèdent une chaîne commune : l'IL-4R α ce qui explique les fonctions partiellement redondantes de ces deux cytokines. Elles permettent la polarisation alternative des macrophages qui produisent alors des médiateurs solubles (e.g. RELM α , ariginase 1) permettant ainsi non seulement la réparation de plaies occasionnées par les parasites mais aussi leur séquestration dans de la matrice extracellulaire. Elles favorisent la contractilité des cellules musculaires lisses et la différenciation des cellules épithéliales en cellules caliciformes productrices de mucus. L'IL-5 favorise quant à elle le recrutement et l'activation des éosinophiles qui vont notamment permettre l'amplification de la réponse immunitaire de type 2 (100). Si ce type de réponse est nécessaire pour lutter contre les helminthes, elle peut également avoir des effets délétères. C'est par exemple le cas dans un contexte d'asthme allergique. Suite à la reconnaissance d'allergènes par les cellules dendritiques, une réponse lymphocytaire de type Th2 se met en place dans les poumons. Elle s'accompagne notamment d'une hyperplasie des cellules caliciformes provoquant une sécrétion excessive de mucus et une bronchoconstriction importante ce qui limite les capacités respiratoires de l'individu et peut avoir des conséquences sévères pour l'organisme (101).

Les gènes *Il4*, *Il13* et *Il5* qui codent pour les cytokines effectrices Th2 sont situés dans une région de 30 KB qui inclut également le gène *Rad50* codant pour la protéine de

réparation de l'ADN RAD50. Ils sont situés sur le chromosome 5 chez l'Homme et sur le chromosome 11 chez la souris. Cette région possède de nombreux enhancers ayant des sites de fixation aux facteurs de transcription canoniques mais également spécifiques des Th2 et a été largement cartographiée dans la littérature. Elle possède également un « locus control region » (LCR) permettant de rapprocher les régions régulatrices entre elles et d'isoler les régions inactives des régions actives notamment. Si toutes les cellules Th2 expriment l'IL-4 il a été montré que seul une fraction des clones de type Th2 produit l'IL-13 et l'IL-5 (102).

c. Les lymphocytes de type Th17

Le lignage Th17 est instruit conjointement par l'IL-6 et le TGF β dont les récepteurs signalent respectivement par STAT-3 et les protéines SMAD. La translocation nucléaire de STAT-3 permet l'expression et la sécrétion d'IL-21 et l'expression de ROR γ (Retinoic acid (RA)-related orphan receptor γ), le « master regulator » du lignage Th17 (103). En synergie avec STAT-3, ROR γ est responsable de la production d'un large panel de cytokines composé notamment de l'IL-17 (IL-17A), de l'IL-17F, de l'hétérodimère IL-17A/F, de l'IL-22, de l'IL-10 ou encore du GM-CSF (104–106). En effet des LT déficients pour ce « master régulateur » ne peuvent produire de l'IL-17 alors que l'expression ectopique de ROR γ en force la production (107). L'expression de l'IL-21 en réponse à l'activation de STAT-3 permet la mise en place d'une régulation positive de ce dernier, et ainsi d'amplifier la différenciation Th17. Les lymphocytes de type Th17 expriment des récepteurs aux chimiokines, notamment CCR6 qui adresse essentiellement les cellules dans la peau et la muqueuse intestinale où l'organisme est souvent confronté à des pathogènes extracellulaires comme des champignons ou certaines bactéries. Les lymphocytes de type Th17 sont en effets spécialisés dans la défense de l'organisme contre ces agents infectieux. L'IL-17A favorise entre-autre la différenciation et le recrutement des neutrophiles sur le site infectieux et l'IL-22 permet par exemple la production de peptides antimicrobiens par les kératinocytes (106, 108).

La présence d'IL-23, composée de la chaîne IL-12p40 (commune à l'IL-12) et de la chaîne IL-12p19, lors de la différenciation des lymphocytes en cellules de type Th17 induit une modulation de leur phénotype. En effet en absence de cette cytokine, les lymphocytes Th17 produisent essentiellement de l'IL-17 et de l'IL-10 (109). La signalisation de l'IL-23R via STAT3 amplifie alors l'expression de l'IL-23R, de ROR γ mais également des cytokines pro-inflammatoires que sont l'IL-22 et le GM-CSF (110). D'autre part elle réprime également la production d'IL-10 (109). Ainsi la réponse Th17 induite en présence d'IL-23 présente un phénotype pro-inflammatoire souvent associé à une réponse pathogénique. Le GM-CSF

notamment joue un rôle prépondérant dans la pathogénicité des lymphocytes de type Th17 (111). Ils ont été mis en cause dans diverses pathologies auto-immunes comme la sclérose en plaque ou dans des pathologies inflammatoires chroniques notamment celle affectant l'intestin.

d. Les lymphocytes T follicular helper (T_{FH})

Contrairement aux autres lignages qui sont caractérisés par leur production de cytokines, les lymphocytes T_{FH} sont essentiellement identifiables par leur sur- et sous-expression d'un ensemble de marqueurs de surface. Ils expriment fortement CXCR5, CD44, ICOS et PD1 mais faiblement CCR7, CD62-L et PSGL1 (CXCR5^{hi} CD44^{hi}, ICOS^{hi}, PD1^{hi}, CCR7^{low}, CD62-L^{low} PSGL1^{low}). CXCR5 est notamment impliqué dans la migration cellulaire et la localisation caractéristique des LT_{FH} dans les zones B notamment les centres germinatifs des OL-II. Et le récepteur PD-1 est un des premiers marqueurs identifiés dont l'expression est fortement corrélée avec la localisation caractéristique des LT_{FH} chez l'homme (112) et la souris (113). Leur localisation est cruciale pour mener à bien leur fonction biologique, puisque les LT_{FH} sont nécessaires à la formation du centre germinatif et à la production d'anticorps de haute affinité par les LB (114).

L'IL-6 et l'IL-21, qui signalent toutes les deux par STAT-3, sont les principales cytokines instructrices de ce type de lymphocytes. La co-stimulation par ICOS-L est aussi nécessaire pour générer des LT_{FH}. En effet, la proportion de cellules T_{FH} est très fortement diminuée en réponse à une immunisation chez des souris déficientes pour l'une ou l'autre de ces protéines (115). Le facteur de transcription BCL-6, essentiellement répresseur, a été identifié comme étant le « master régulateur » du lignage T_{FH}. En effet, les LT_{FH} sont absents chez les souris *Bcl6*^{-/-} qui sont dépourvues de centres germinatifs. De plus l'expression ectopique de BCL-6 dans des lymphocytes Th1 et Th17 induit l'acquisition de CXCR5 par ces cellules et leur migration dans les centres germinatifs (116, 117).

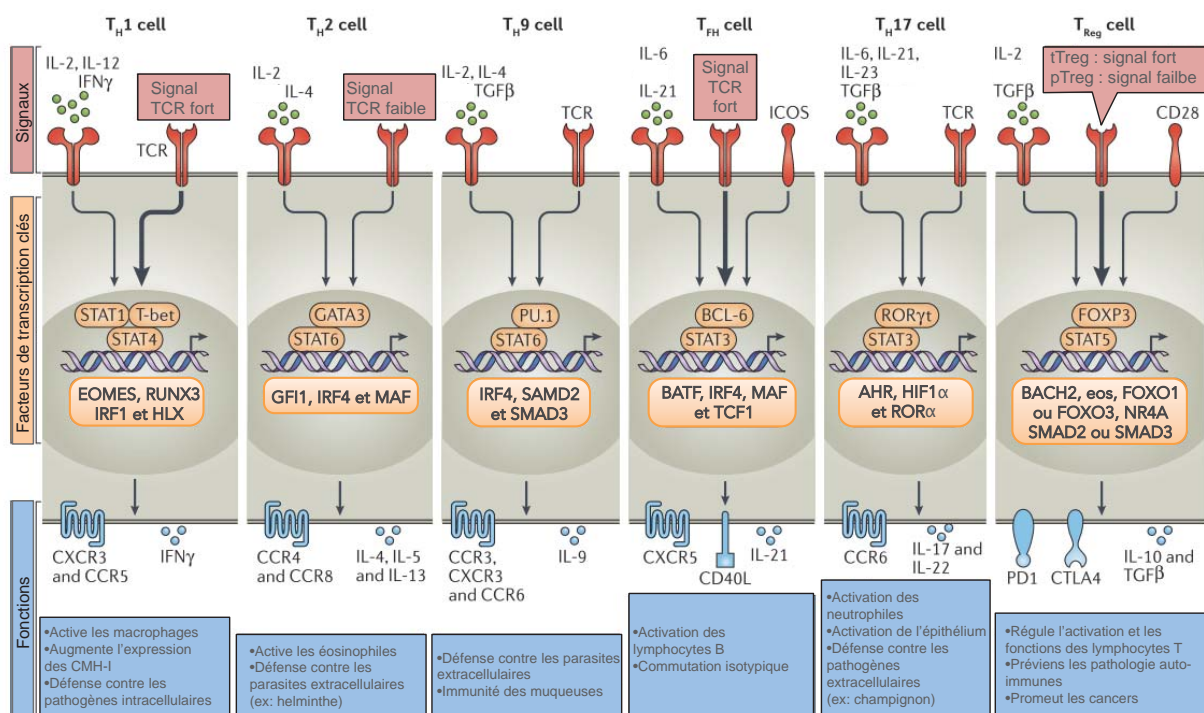


Figure 7 : Différents signaux environnementaux instruisent la différenciation d'un large répertoire de lignages T CD4⁺ caractérisés par des signatures moléculaires et des fonctions distinctes

Les signaux extracellulaires nécessaires à l'engagement et à la différenciation des LT_N dans les différents lignages T helper sont décrits en rouge. Ces signaux, essentiellement délivrés par le TCR et les récepteurs aux cytokines, initient la mise en place d'un réseau de facteurs de transcription coordonné notamment par des facteurs pionniers (e.g. STAT) et les « master régulateurs » de chaque lignage (en orange). Enfin, les principales molécules effectrices et leurs impacts sur la réponse immunitaire sont exposés pour chaque sous-population (en bleu). Adapté de (82)

2. Du paradigme Th1/2... au continuum de différenciation

Le paradigme Th1/Th2 est proposé par Mosmann et Coffman en 1986 (118). Dans cet article princeps, ces auteurs ont mis en évidence que les clones de lymphocyte T helper pouvaient être ségrégués en deux catégories selon le répertoire de « lymphokines » qu'ils produisaient. Quelques années plus tard, le groupe de Richard Locksley établit que ces deux populations exercent des fonctions biologiques distinctes. Alors que les LT Th1 assurent la clairance du parasite *Leishmania major* chez la souris, les cellules Th2 sont incapables d'éradiquer le pathogène après transfert adoptif dans des animaux lymphopéniques infectés (119). Le paradigme Th1/Th2 reposait donc sur le fait que les LT_N étaient capables de se différencier en deux sous-populations aux caractéristiques fonctionnelles distinctes. Ce paradigme a été renforcé par la mise en évidence de la stabilité et de l'antagonisme de ces lignages. En effet le groupe de Richard Locksley a montré que des LT de type Th1 ou Th2 ayant effectué au moins quatre cycles cellulaires étaient pleinement déterminés dans leur lignage et ne pouvaient plus acquérir les caractéristiques phénotypiques et fonctionnelles de l'autre sous-population, même en présence de forte dose de signal instructeur (69). Cette perte de potentialité était en partie expliquée par la perte d'expression, au cours de la différenciation, des récepteurs aux cytokines instructrices du lignage opposé. Elle découlait aussi de l'antagonisme existant entre T-bet et GATA-3 (120, 121). Ces travaux initiaux, associés à de nombreux autres par la suite, ont abouti pendant près de vingt ans à l'idée que les lymphocytes Th1 et Th2 constituaient des lignages distincts, en « stade terminal de différenciation ».

Peu de temps après l'établissement du paradigme, il a été mis en évidence par Manetti et collaborateur que l'IL-12, en plus de favoriser la différenciation Th1, inhibait la production d'IL-4 (122). Si les événements moléculaires sous-jacents à cette observation n'avaient pas été identifiés dans cette étude, plusieurs mécanismes responsables de la répression du programme Th2 dans cellules Th1 ont depuis été documentés. Lors de la différenciation Th1, Tbet est rapidement phosphorylé par la kinase ITK sur la tyrosine 525. La forme phosphorylée de Tbet peut alors interagir physiquement avec GATA-3 et l'empêcher de se fixer sur l'ADN pour induire la transcription de ses gènes cibles, notamment celui de l'IL-5 (120). De plus, il a été montré que l'expression ectopique de Tbet dans les cellules Th2 en cours de différenciation diminuait l'expression de GATA-3 et des cytokines Th2 (121). Enfin, Tbet est capable de réprimer directement l'expression de l'IL4. Il coopère en effet avec le facteur de transcription RUNX3 pour activer le *silencer* HS IV (DNaseI Hypersensitivity Site IV) localisé en aval du gène. Il s'agit ici d'un mécanisme important de répression d'une cytokine antagoniste

car les LT Th1 *Runx3*^{-/-} n'ont pas la capacité de réprimer l'*Il4* et produisent cette cytokine de façon aberrante (84). Enfin, il a été suggéré que le recrutement des récepteurs aux cytokines au niveau de la synapse immunologique pourrait être mutuellement exclusif, empêchant ainsi les cellules de percevoir des signaux alternatifs. Le recrutement du récepteur à l'IFN- γ (IFNGR) au niveau de la synapse immunologique prévient par exemple celui de l'IL-4R qui est alors exclu en périphérie de celle-ci (53). Néanmoins, cela présuppose que les LT ne captent les cytokines qu'au niveau de la synapse ou que les récepteurs aux cytokines ne peuvent signaler que dans cette zone cellulaire, ce qui n'a pas été démontré de façon convaincante.

Les LT de type Th1 répriment le programme Th2 mais la réciproque est également vraie. GATA-3 peut en effet inhiber le programme d'expression génique induit par l'IL-12 dans les cellules Th1. L'expression ectopique de ce facteur de transcription dans des cellules Th1 en cours de différenciation inhibe la production d'IFN- γ (123) en réprimant l'induction de l'IL-12R β 2 (124). GATA-3 joue donc un rôle précoce dans l'exclusion mutuelle des deux lignages. De plus il a été montré, dans des LT exprimant de façon constitutive l'IL-12R β 2 et pouvant donc se différencier correctement en LT de type Th1, que l'expression ectopique de GATA-3 réprimait la production d'IFN- γ en inhibant l'expression de STAT-4 (124). Enfin, il a été montré que l'activation de STAT-6 induite par l'IL-4 prévenait le recrutement de l'IFNGR au niveau de la synapse immunologique (53).

Enfin, les huit protéines SOCS (Suppressor Of Cytokine Signaling) jouent un rôle primordial dans la stabilité des LT helper. Les SOCS se lient aux récepteurs aux cytokines et aux kinases JAK pour induire leur dégradation, respectivement par la voie lysosomale et le protéasome. L'action combinée des différentes SOCS permet non seulement d'exercer un contrôle négatif de la signalisation des STATs au cours de la différenciation des LT helper, mais également de limiter les risques de repolarisation ou transdifférenciation. Ainsi, SOCS5 diminue la signalisation de STAT6 dans les cellules Th2 en cours de différenciation, et SOCS3 fait de même dans les cellules Th1 en inhibant la signalisation de STAT-4 en aval de l'IL-12R (125).

Pendant deux décennies, l'idée que seules ces deux populations de LT helper existaient a prévalu dans la communauté des immunologistes. Les réponses lymphocytaires étaient classifiées comme étant de type Th1 ou Th2 sur la seule base de l'expression par les cellules T CD4 des facteurs de transcription T-bet et GATA-3, et de leur sécrétion d'IFN- γ et IL-4. En 1995, ce paradigme fut questionné une première fois. Il fut en effet proposé que les LT Th1 et les LT Th2 ne représenteraient que deux extrémités d'un continuum de différenciation des LT_H (126).

Si dès 1995, Shimon Sakaguchi révèle dans son article fondateur l'existence d'une population de lymphocytes T CD4 régulateurs (127), la diversité des sous-types de LT helper n'a été appréhendée que dans les années 2000 avec la caractérisation des cellules T_{FH} (114) et Th17 (104, 105). Plus récemment, des études ont également mis en évidence des LT producteurs d'IL-9 : les LT de type Th9. Ils peuvent être générés *in vitro* à partir de LT_N cultivés en présence de TGF- β et d'IL-4 (128), ou à partir de LT de type Th2 également cultivés en présence de TGF- β qui sont ainsi reprogrammés pour produire de l'IL-9 (129). La communauté scientifique s'est par la suite interrogée sur une potentielle plasticité des LT effecteurs, c'est-à-dire sur leur capacité à adapter leur phénotype à l'environnement.

Les lymphocytes T de type Th17 constituent une population cellulaire très plastique. L'utilisation d'un modèle murin permettant de marquer de façon « indélébile » les cellules produisant ou ayant produit de l'IL-17, a notamment permis de montrer que les cellules Th17 peuvent acquérir un phénotype régulateur, et qu'elles participent alors de façon critique à la résolution de l'inflammation. Ces LT, qualifiés d'« exTh17 », n'expriment pas Foxp3 mais acquièrent un phénotype Tr1. Leur fonctions immunosuppressives et anti-inflammatoires reposent en effet essentiellement sur leur capacité à produire de l'IL-10 (130). La présence de ces lymphocytes exTh17 a notamment été révélée dans un modèle murin de sclérose en plaque et suite à l'infection d'animaux de laboratoire avec des helminthes. Les cellules Th17 sont également plastiques vers le lignage Th1. En effet il apparaît qu'après plusieurs cycles de culture en conditions pro-Th17, les cellules Th17 expriment faiblement l'IL-12 α 2, la chaîne inductible du récepteur à la cytokine instructrice du lignage Th1, et maintiennent de ce fait la capacité de produire l'IFN- γ . Ainsi, lorsqu'elles sont restimulées en présence d'IL-12, les cellules Th17 produisent de l'IFN- γ de façon dépendante de Tbet et STAT4 (131). Cette plasticité vers le phénotype Th1 a également été observée *in vivo* et a été associée à la pathogénicité des cellules Th17. En effet, contrairement à leurs contreparties sauvages, le transfert adoptif de cellules Th17 *Ifng*^{-/-} chez des souris lymphopéniques ne permet pas d'induire une colite expérimentale (132). Enfin, chez l'Homme comme chez la souris, des LT présentant une signature moléculaire mixte Th17 et Th2 ont pu être identifiés chez des individus atteints d'inflammation allergique des voies respiratoires (133–135).

Les cellules T_{FH} sont également très plastiques puisqu'elles peuvent acquérir des sous-modules de programmation appartenant aux autres sous-populations T helper ou au lignage régulateur Foxp3⁺ (136). Ainsi des LT T_{FH} cultivés *in vitro* en présence des cytokines polarisantes d'autres types de lymphocytes T helper acquièrent un phénotype mixte à la fois caractéristique du type T_{FH} et de type alternatif. Ce phénotype mixte a un intérêt biologique. Il

a en effet été montré, grâce à l'infection par *Leishmania major* d'un modèle murin double rapporteur pour l'IFN- γ et l'IL-4, que des LT T_{FH} de type mixte Th1 ou Th2 permettaient la commutation isotypique des anticorps en IgG_{2a} et IgG1, respectivement (137). Ainsi les cytokines des types alternatifs produites par les LT de type T_{FH} au niveau des synapses LT-LB façonnent le répertoire d'immunoglobulines produites par les LB afin de les adapter au danger.

Concernant les LT régulateurs, plusieurs études ont rapporté que ceux-ci perdaient leur stabilité dans des environnements pro-inflammatoires. Les LT régulateurs expriment alors simultanément leur master régulateur, Foxp3, et celui d'un type de lymphocytes T helper alternatif tel que T-bet, GATA-3 ou BCL-6 (138–141). L'intérêt biologique de cette plasticité serait de réguler les réponses Th1, Th2 et T_{FH} respectivement. Mais cette théorie reste controversée (142, 143). Une étude a notamment montré que l'acquisition de T-bet par les LT Treg était associée aux décès des souris infectées par *Toxoplasma gondii* (143).

Les lignages Th1 et Th2 restent quant à eux les plus stables identifiés à ce jour et sont parfois considérés comme des stades terminaux de différenciation des LT CD4. Néanmoins la possibilité de reprogrammer les lymphocytes Th2 en cellules Th9 a remis cette notion en cause. Cependant, contrairement aux lignages TH17 ou T_{FH} très plastiques, peu de publications font état de la plasticité du lignage Th1 vers le lignage Th2 ou inversement. Le groupe de Max Löhning fait exception en montrant que des LT de type Th2 spécifiques du LCMV et générés *in vitro* co-exprimaient les facteurs de transcription GATA-3 et T-bet et produisaient simultanément de l'IL-4 et de l'IFN- γ suite à leur transfert adoptif dans des hôtes infectés par le LCMV, un virus impliqué dans des réponses pro-Th1 (144). Les LT de type mixte Th2+1 GATA-3⁺Tbet⁺IL-4⁺IFN- γ ⁺ ainsi générés perdurent jusqu'à soixante jours après infection témoignant d'une relative stabilité de ce phénotype mixte. Les auteurs n'expliquent pas les mécanismes moléculaires évitant la répression mutuelle des programmes Th1 et Th2 mais ce phénotype mixte peut être mimé *in vitro* par la culture des Th2 spécifiques du LCMV en présence d'IL-12, d'IFN- γ et d'IFN de type I ou la culture de LT_N en présence d'IL-4, d'IL-12 et d'IFN- γ (144, 145). *In vivo*, des LT de type mixte sont spontanément générés dans des modèles infectieux à *Heligmosomoides polygyrus* et *Schistosoma mansoni* de façon dépendante de la signalisation de l'IFN- γ (145, 146). Pour preuve, les souris déficientes pour *Ifngr1* ne développent pas de LT de type mixte Th2+1 après infection par *H. polygyrus*.

La proportion croissante d'études révélant l'hétérogénéité et la plasticité des LT remet donc en question, au moins en partie, le paradigme Th1/Th2 en faveur de la différenciation des LT selon un spectre continu dépendant de l'exposition des cellules à des combinaisons

Les ellipses de couleurs représentent les zones d'influence des cytokines qui contrôlent le devenir des LT CD4⁺. Les LT_N sont des cellules hautement plastiques qui peuvent se différencier en une multitude de lignages T helper. Selon leur degré de détermination (représenté ici de façon inversement corrélé à la taille des cellules) et la combinaison de cytokines présentes dans leur microenvironnement, les phénotypes acquis peuvent être stables et spécifiques d'un lignage ou au contraire mixtes et relativement plastiques. Adaptée de (82).

Chapitre II : Régulation de l'expression des gènes

I. Organisation de la chromatine

Le génome d'une cellule humaine mesure environ 2 mètres de long. En dehors de l'ADN mitochondrial, il est organisé et compacté dans des noyaux cellulaires de quelques micromètres de diamètre par de petites protéines basiques, nommées histones, d'environ 15 kDa de masse moléculaire et très conservées au cours de l'évolution. L'ensemble ADN et histones forme la chromatine dont l'unité de base est le nucléosome (**Figure 9**). Deux copies de chacune des histones H2A, H2B, H3 et H4 s'associent pour former un octamère qui constitue le cœur protéique du nucléosome autour duquel s'enroulent 146 paires de bases d'ADN (147). Ces histones de cœur ont une structure similaire. Leur domaine carboxy-terminal globulaire est composé d'un motif « histone-fold » comprenant trois hélices α séparées par deux boucles. Il permet leur hétérodimérisation en formant un motif d'interaction dit « en poignée de main ». Leur extrémité amino-terminale, plus variable, est quant à elle dépourvue de structure secondaire. Elle est très basique car riche en résidus lysine et arginine. Exposée à la surface du nucléosome, elle est la cible de nombreuses modifications post-traductionnelles covalentes qui affecteront notamment la charge des histones et les interactions protéine/protéine et protéine/ADN. Le nucléosome est finalement complété et stabilisé par l'histone de liaison H1 qui se positionne à l'endroit où le filament d'ADN entre et sort des particules de cœur.



Figure 9 : Constitution du nucléosome

Lors de l'assemblage de la chromatine, un tétramère d'histones $(H3-H4)_2$ forme une particule sub-nucléosomale sur laquelle deux dimères H2A-H2B viennent s'arrimer pour former un octamère autour duquel l'ADN s'entoure pour former le cœur du nucléosome. (148)

Les nucléosomes sont reliés entre eux par des fragments d'ADN de 20 à 80 paires de bases pour former une structure dite « en collier de perle » de 11 nm de diamètre (**Figure 10**). Ces filaments s'organisent et se compactent pour constituer une fibre de chromatine de 30 nm qui forme des boucles d'environ 300 nm de large. Celles-ci peuvent elles-mêmes se condenser pour former des structures encore plus complexes dont l'organisation tridimensionnelle reste à ce jour matière à débat.

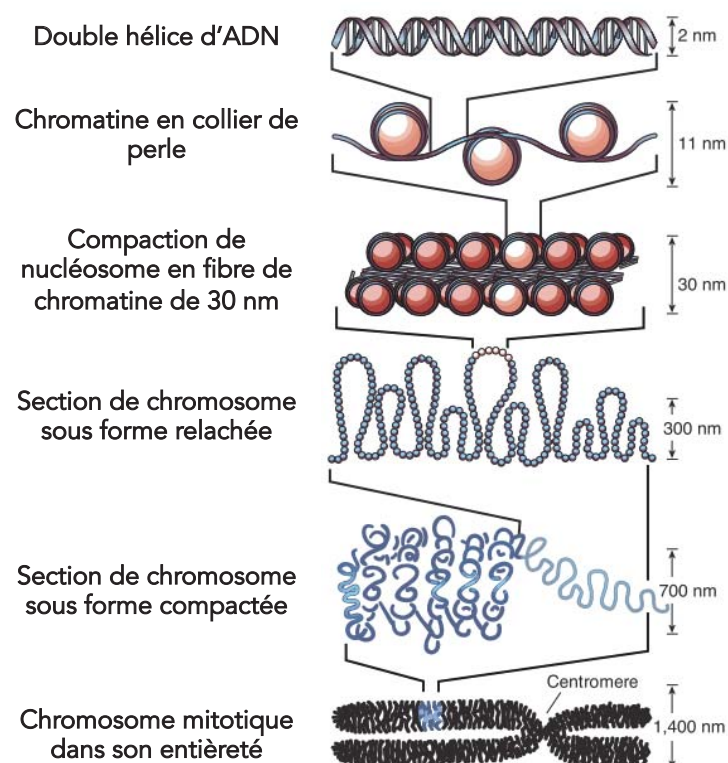


Figure 10 : Organisation de la chromatine

La double hélice d'ADN s'organise autour d'un octamère d'histones pour former le nucléosome. Celui-ci est stabilisé par l'histone de liaison H1 qui interagit avec les histones du cœur et l'ADN de liaison entrant et sortant du nucléosome. Le collier de perle ainsi formé (11nm) est davantage compacté en fibre de 30 nm formant elle-même des boucles de 300 nm. Un niveau de compaction supplémentaire de la chromatine aboutit aux chromosomes mitotiques tels qu'ils sont observés sur un caryotype. Adapté de (148)

Dans le noyau des cellules eucaryotes, la chromatine est présente de façon schématique sous deux formes distinctes, l'euchromatine et l'hétérochromatine. L'euchromatine est répartie à l'intérieur du nucléoplasme. Elle est faiblement condensée puisqu'elle correspond à des fibres de 30 nm ou moins. Elle apparaît donc peu dense aux électrons en microscopie

électronique (**Figure 11**). Elle est riche en gènes, notamment codants, et a une activité transcriptionnelle élevée. Sa décompaction pendant l'interphase permet en effet le recrutement et l'assemblage de la machinerie de transcription. Les régions du génome sous forme d'euchromatine diffèrent selon le type cellulaire considéré et selon son stade d'activation ou de différenciation. Des régions ouvertes peuvent en effet se fermer et se condenser sous forme d'hétérochromatine facultative si les gènes qu'elles contiennent ne sont plus nécessaires aux fonctions de la cellule. A l'inverse, une région du génome verrouillée sous forme d'hétérochromatine facultative peut s'ouvrir si la cellule reçoit un signal instruisant la mobilisation et l'expression des gènes qu'elle contient (149). Les états d'euchromatine et d'hétérochromatine facultative sont donc réversibles et certains *loci* peuvent passer de façon dynamique d'un état à l'autre en fonction des signaux intégrés par la cellule. L'hétérochromatine existe sous une autre forme, qualifiée de constitutive. Contrairement aux autres états de la chromatine, l'hétérochromatine constitutive est stable. Elle garde ses propriétés dans tous les types cellulaires, quel que soit le tissu ou le stade de différenciation. Elle apparaît en périphérie du noyau et dense aux électrons, car très compacte, en microscopie électronique. Elle est essentiellement localisée au niveau des régions péricentromériques et télomériques, ce qui explique qu'elle soit dénuée en gène et riche en ADN satellite (séquences courtes et répétées en tandem un très grand nombre de fois). D'un point de vue fonctionnel, l'hétérochromatine constitutive a un rôle critique dans la stabilité et l'intégrité des chromosomes notamment au niveau des télomères, la cohésion des chromatides sœurs et la ségrégation des chromosomes lors de la mitose (150, 151).

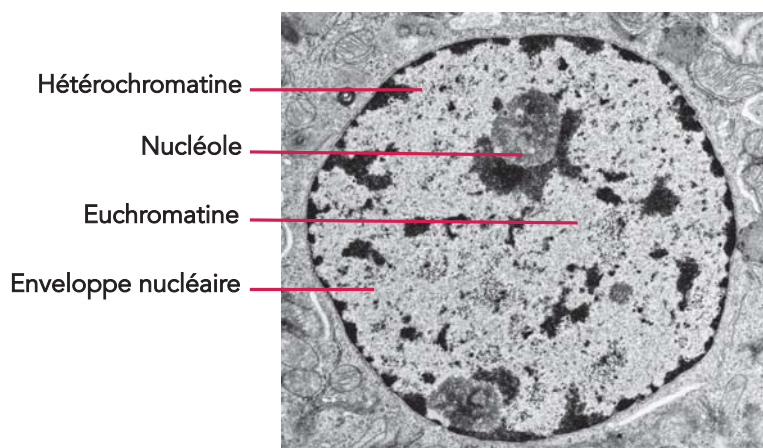


Figure 11 : Compaction de la chromatine

Noyau cellulaire en interphase vu en microscopie électronique. L'hétérochromatine (noire) est dense aux électrons de par sa structure condensée. L'euchromatine (claire) est moins condensée et transcriptionnellement plus active. Adapté de <http://medcell.med.yale.edu>.

Dans le noyau des cellules eucaryotes, le filament d'ADN, support de l'information génétique, s'associe donc aux protéines histones pour former la chromatine dont le niveau de compaction va être modulé dans le temps et l'espace par les signaux environnementaux, permettant ainsi à la cellule d'ajuster finement son programme d'expression génique. Parmi les mécanismes épigénétiques qui contrôlent l'état et la composition de la chromatine, et donc l'accessibilité à l'information génique, la méthylation de l'ADN et les modifications post-traductionnelles des histones jouent un rôle majeur.

II. Eléments *cis*-régulateurs de l'expression génique

Le niveau de compaction de la chromatine au niveau d'un gène est essentiel en lui-même pour réguler la transcription génique. Cependant l'activation d'un gène requiert également la coordination de nombreux facteurs au niveau de son promoteur et de régions plus distantes qualifiées de « *cis*-régulatrices » car situées à proximité du gène cible (**Figure 12**).

La première région « *cis*-régulatrices » à avoir été décrite est le promoteur des gènes qui agit comme un interrupteur pour activer la transcription. Il est tout d'abord constitué d'un cœur (« core ») qui correspond à la séquence d'ADN minimale requise pour initier la transcription. Il sert de plateforme d'assemblage au complexe d'initiation de la transcription (152). Cette région d'environ 100 pb inclut le site d'initiation de la transcription (TSS pour Transcription Start Site), et permet le recrutement de l'ARN polymérase II et de facteurs généraux de la transcription (TFII). La composition en motifs de fixation pour ces facteurs est variable. Un des plus remarquables est la « TATA box » et tire son nom de la richesse de sa séquence en adénine et thymine. Il permet le recrutement de la sous-unité « TATA-binding protein » (TBP) du facteur TFIIID, important pour l'initiation de la transcription. Parmi les autres motifs connus, on retrouve l'élément de reconnaissance par le facteur TFIIIB (« BRE ») et l'initiateur (« Initiator » ou *Inr*) dont la fonction est similaire à la TATA box. La machinerie basale de transcription est ainsi recrutée et assemblée au niveau du cœur du promoteur. Elle implique une interaction entre la chromatine, des protéines régulatrices et la machinerie générale de transcription constituée de la RNA pol II et des facteurs TFIIA-B-D-E-F et H. Le promoteur comprend également le promoteur « proximal ». Sa séquence se situe juste en amont du « core » promoteur et mesure environ 250 pb. Il contient divers sites de fixation pour des facteurs de transcription dont la liaison module le niveau de la transcription. Enfin il peut exister un

promoteur plus distal contenant des sites supplémentaires pour la fixation d'éléments régulateurs.

La régulation de l'expression d'un gène est complexe et après l'activation, le niveau d'expression d'un gène est également modulé par des régions plus distales comme les activateurs transcriptionnels aussi appelés « enhancers ». Il s'agit de régions amplificatrices de l'expression génique, localisés de façon intragéniques ou jusqu'à 1Mb du TSS du gène dont ils régulent l'expression. Ils contiennent des sites de fixation pour des facteurs de transcription. Les enhancers peuvent agir sur leur promoteur cible par différents mécanismes, le plus connu étant la formation d'une boucle chromatinienne qui permet de rapprocher physiquement ces deux éléments comme cela a été montré dans la régulation des gènes *Il4*, *Il5*, *IL13* et *Rad50* du locus *Th2* (153). Un même gène peut être associé à plusieurs enhancers qui seront utilisés ou non par la cellule en fonction de son type et de son stade de différenciation. On peut enfin retrouver des clusters d'enhancers fortement actifs, on parle alors de super-enhancers (154). Ils sont caractérisés par un fort enrichissement en histone acétyltransférase p300 ou en histone H3 acétylée sur la lysine 27 (H3K27ac). Les super-enhancers sont liés par un grand nombre de co-activateurs transcriptionnels, l'ARN polymérase II et des complexes de remodelage de la chromatine. Ils contrôlent ainsi l'expression de gènes nécessaires au déterminisme cellulaire.

Il existe d'autres éléments *cis*-régulateurs dont font partie les régions inhibitrices de l'expression des gènes, ou « silencers ». Elles sont liées par des facteurs de transcription spécifiques qui répriment l'expression des gènes. Ces régions sont encore relativement peu étudiées. Néanmoins leurs mécanismes d'action pourraient reposer soit sur la propagation locale d'hétérochromatine, soit sur la formation de boucles de chromatine qui rapprocheraient les facteurs répresseurs liés aux « silencers » et les promoteurs (155). Les facteurs répresseurs pourraient alors interférer avec la liaison des facteurs de transcription généraux nécessaires à l'expression génique. Enfin la dernière catégorie d'éléments *cis*-régulateurs notables correspond à des éléments de délimitation également appelés « insulators » ou « boundaries ». Ils bloquent l'action d'enhancers ou de répresseurs transcriptionnels en interrompant leur interaction avec le promoteur d'un gène via la formation de boucle de chromatine (156). Les mécanismes moléculaires sous-jacents restent cependant flous chez les mammifères. Néanmoins plusieurs études ont révélé que le facteur à doigt de zinc CTCT peut se lier à la plupart des « insulators » et induire des boucles de chromatine longue distance.

La composition et le niveau de compaction de la chromatine au niveau de ces séquences régulatrices a un impact considérable sur leur capacité à recruter leurs partenaires moléculaires. Leur activité est donc sous le contrôle strict de voies de régulation épigénétiques

dont les interactions et les principaux acteurs moléculaires restent encore largement à identifier.

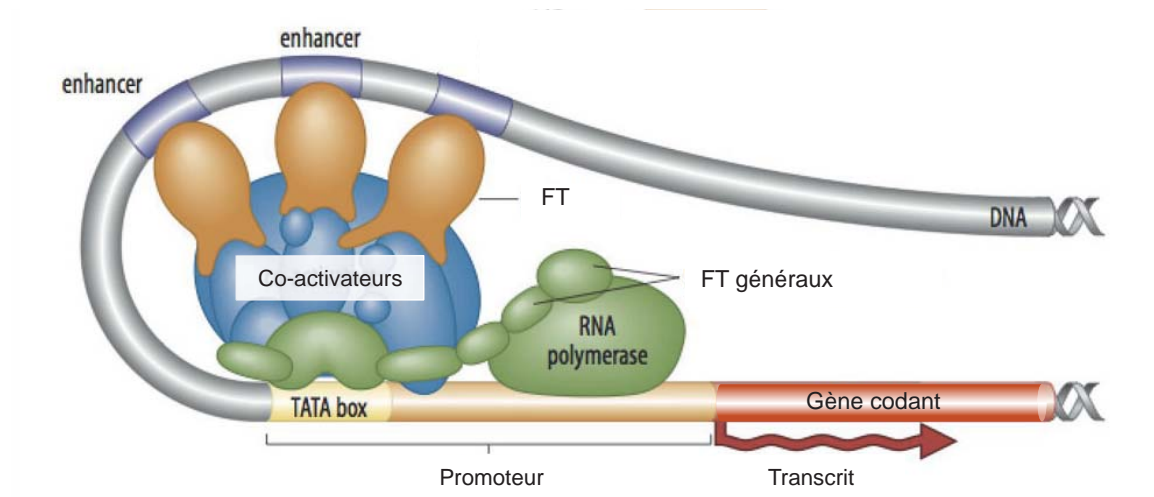


Figure 12 : Éléments *cis*-régulateurs de l'expression génique

Représentation schématique d'éléments *cis*-régulateurs de l'expression génique. Des facteurs de transcription généraux se fixent indirectement ou non sur la TATA box et recrutent l'ARN polymérase II. D'autres facteurs de transcription reconnaissent des séquences enhancer à distance du promoteur et permettent d'amplifier la transcription génique. FT, Facteur de transcription. Adapté de <http://www.mun.ca>.

III. Régulation épigénétique

A. Historique et concept

A de rares exceptions près, dont font partie les lymphocytes B et T puisqu'ils ont réarrangé les loci codant pour leur récepteur à l'antigène, toutes les cellules d'un organisme possèdent le même génome nucléaire. Comment expliquer alors, qu'à partir d'une seule et même partition, une multitude de types cellulaires aux fonctions et phénotypes extrêmement différents puissent émerger ? Pour aborder cette question, Conrad H. Waddington, considéré comme le père de l'épigénétique moderne, « fusionna » deux domaines de recherche : la biologie du développement et la génétique. Contrairement à la plupart de ses pairs qui défendaient alors que les gènes ne contrôlent que des phénotypes mineurs, il postula dès les années 30 que l'embryogénèse était un processus coordonné au niveau génétique. En 1942, il définit l'épigénétique comme la science qui s'intéresse aux mécanismes qui contrôlent le génotype afin d'aboutir à l'expression d'un phénotype (157). Il illustra alors sa pensée à l'aide d'un dessin : « le paysage épigénétique » (**Figure 13**). Il représente une pierre ronde en haut d'un paysage vallonné lui offrant plusieurs chemins possibles pour dévaler la pente. Cette pierre est en fait l'illustration d'une cellule multipotente et les différentes vallées représentent les devenirs qui lui sont offerts. Au fur et à mesure de sa descente, la cellule se différencie et se détermine toujours plus dans un lignage donné par les « choix » qu'elle réalise à chaque intersection. Le relief étant façonné par des réseaux fonctionnels de gènes, il illustre également le niveau de tolérance des phénotypes cellulaires vis-à-vis des fluctuations de l'expression des gènes et de l'environnement. Ainsi, un relief marqué entre deux vallées illustre la capacité des cellules à rester stable malgré les fluctuations. Cette capacité d'une cellule ou d'un organisme à conserver ou reproduire un phénotype dans un environnement changeant était définie par Waddington par le terme de « canalisation ». Ce modèle n'excluait pour autant pas la possibilité pour le vivant de s'adapter à son environnement puisque Waddington avait aussi documenté l'existence d'une « assimilation génétique », c'est-à-dire la capacité d'un organisme à encoder de façon stable dans son (épi-)génome un phénotype initialement induit en réponse à un stress environnemental. Même si le processus décrit par Waddington n'était au départ pas Lamarckien, dans le sens où il ne démontrait pas l'héritabilité des caractères acquis, Robin Holliday et Art Riggs complèteront 20 ans plus tard la définition de l'épigénétique avec la notion d'héritabilité (158, 159). En effet, ils avaient observé des changements héréditaires dans l'expression des gènes après méiose ou mitose sans qu'il n'y ait d'altération de la séquence d'ADN.

L'épigénétique devient alors l'étude des mécanismes transmissibles, à l'échelle cellulaire ou de l'organisme, qui régulent la fonction du génome sans altération de la séquence de base de l'ADN. Nous savons aujourd'hui que ces mécanismes mobilisés en réponse aux signaux de l'environnement, s'ils sont héréditaires, sont aussi réversibles.

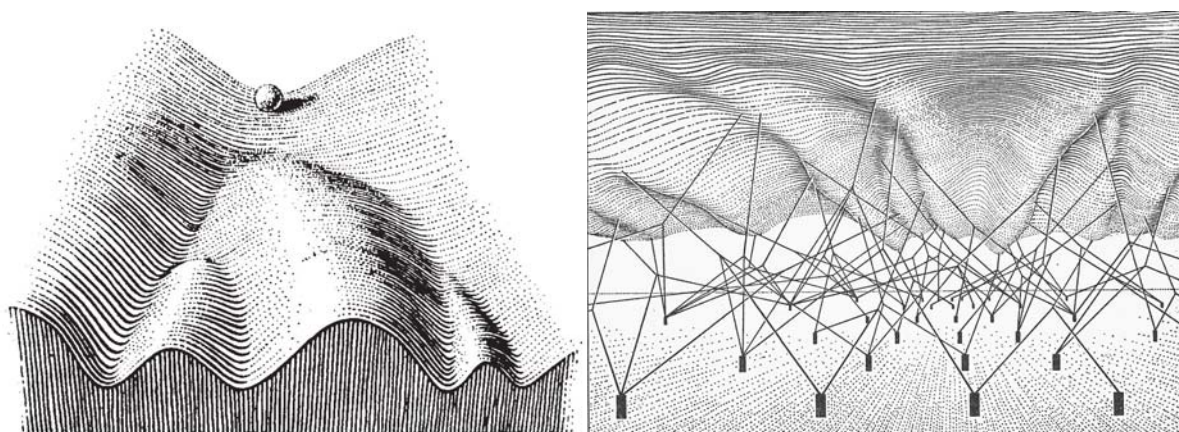


Figure 13 : Le paysage épigénétique de Conrad Waddington

Il s'agit d'une métaphore visuelle du développement d'un organisme complexe ou plus simplement de la différenciation cellulaire. (Gauche) Le relief représente les choix successifs qu'une cellule (la bille) doit réaliser au cours de sa différenciation. Ces choix sont influencés par l'environnement et aboutissent à la détermination cellulaire c'est-à-dire à la spécialisation de la cellule mais aussi à une perte de potentialité. (Droite) Le paysage est en partie façonné par des forces exercées par des réseaux de gènes, illustrés ici par des boîtes noires (gènes) reliées par des cordes. Issus de « The Strategy of genes » de C. H. Waddington (1957).

Les modifications épigénétiques correspondent à l'ajout d'un groupement biochimique, qualifié de marque ou de tag épigénétique. Ces marques sont générées par un ensemble hétérogène d'enzymes pouvant être regroupées sous le terme générique de « chromatin writers ». Les marques qu'elles déposent, que ce soit sur l'ADN ou sur les protéines chromatinienne, peuvent être lues par des « chromatin readers » (**Figure 14**). Même si des marques telles que l'acétylation des histones ont un impact direct sur la fonction de la chromatine, la majorité d'entre-elles n'a pas par elle-même d'effet biologique. Le recrutement de « chromatin readers » initie l'assemblage de complexes protéiques multimériques, et ce sont les molécules ainsi recrutées, en modifiant la composition et la structure de la chromatine, qui sont responsables de l'effet biologique associé à la modification. Les marques épigénétiques, si elles sont transmissibles d'une cellule mère à une cellule fille, sont de plus

dynamiques et potentiellement réversibles. Leur disparition peut être liée à une absence de déposition sur le brin fille lors de la réplication de l'ADN, mais elle peut aussi résulter de l'action d'enzymes que l'on regroupe sous le terme de « chromatin erasers ». L'action coordonnée de ces trois catégories de protéines permet d'écrire l'épigénome d'une cellule, de le traduire en effet biologique, et enfin d'adapter la partition épigénétique à l'environnement.

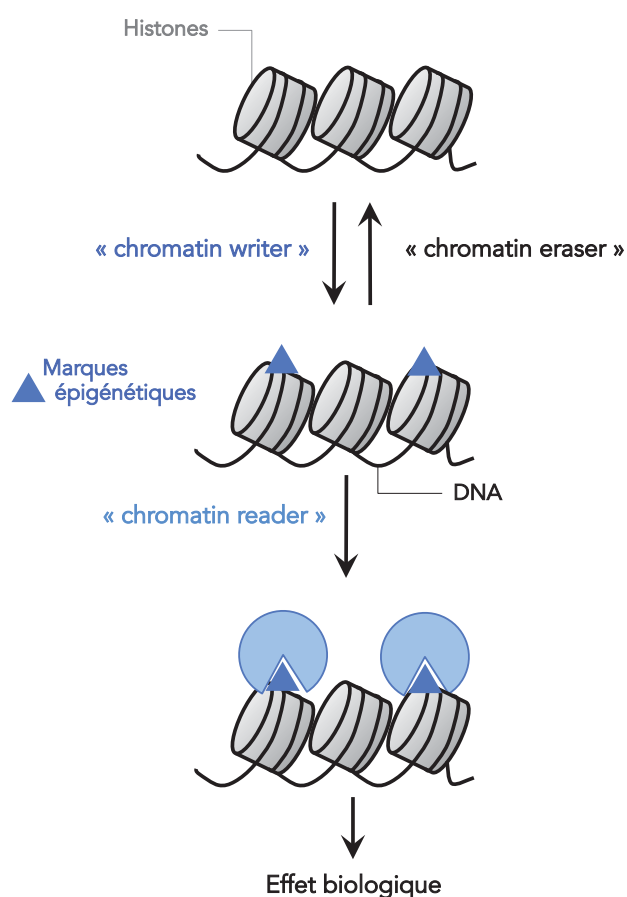


Figure 14 : "Chromatin writer", "reader" et "eraser"

Les différentes marques épigénétiques sont déposées par des enzymes qui peuvent être regroupées sous le nom générique de "chromatin writers". Ces marques sont lues par des protéines, appelées "chromatin readers", très souvent responsables des effets biologiques qui leur sont associées. La déposition des marques n'est pas irréversible puisqu'elles peuvent être activement enlevées par des "chromatin erasers".

Parmi les mécanismes épigénétiques qui contrôlent l'état et la composition de la chromatine, la méthylation de l'ADN et les modifications post-traductionnelles des histones jouent un rôle majeur. Je focaliserai donc la suite de mon manuscrit sur ces deux types de processus.

B. Méthylation de l'ADN

Si les mécanismes épigénétiques n'altèrent pas la séquence d'ADN, ils peuvent cependant cibler des bases azotées afin de les modifier biochimiquement. Chez les procaryotes, les adénines et les cytosines peuvent être méthylées. La méthylation des adénines, bien qu'étant fréquente au niveau des ARN, ne semble par contre pas jouer de rôle majeur au niveau de l'ADN chez les eucaryotes (160). En revanche, la méthylation du carbone 5 de l'anneau pyrimidique des cytosines (5mC) est une modification épigénétique fréquente. Elle affecte principalement les cytosines présentes au niveau des di-nucléotides CpG, mais peut aussi marginalement intervenir hors de ce contexte (161–163). Contrairement aux procaryotes, l'ADN des eucaryotes est globalement hyperméthylé. Le taux de méthylation peut en effet atteindre 80% au niveau des di-nucléotides CpG dont la distribution n'est pas homogène à travers le génome. Les di-nucléotides CpG contenus dans des régions non soumises à une pression de sélection ont en effet été convertis au cours de l'évolution par déamination des 5-méthylcytosines en thymidine. Ce processus a abouti à une sous-représentation des di-nucléotides CpG dans le génome puisqu'ils n'ont essentiellement été conservés qu'au niveau du corps des gènes, de leurs régions promotrices ou enhancers. Ils s'accumulent notamment au niveau des promoteurs des gènes de ménage (164) et sont alors qualifiés d'« îlots CpG ». Les gènes de ménage étant exprimés de façon constitutive, et la méthylation des cytosines étant essentiellement associée à de la répression génique, ces îlots CpG sont souvent non méthylés (165). Au niveau des régions cis-régulatrices et du corps des autres gènes, la méthylation des cytosines est retrouvée de façon spécifique du type cellulaire, selon que la cellule exprime ou non les gènes. Enfin, l'ADN satellite et les éléments répétés au niveau de l'hétérochromatine sont aussi hautement méthylés.

Comme toute marque épigénétique, la méthylation des cytosines est catalysée par des « chromatin writers », les ADN méthyl-transférases (DNMT). Elle peut être lue par des « chromatin readers » se liant au 5mC (MeCP) qui recruteront d'autres partenaires. Enfin s'il est maintenant admis qu'elle peut être effacée par des mécanismes actifs, il n'existe pas de véritable « chromatin eraser ». La déméthylation de l'ADN dépend en effet de cascades enzymatiques complexes faisant notamment intervenir les protéines TET et le complexe de réparation de l'ADN BER.

1. ADN méthyl transférases (DNMT)

Trois membres de la famille des DNMT peuvent transférer un groupement méthyl du cofacteur S-adénosyl méthionine (SAM) sur le carbone 5 de l'anneau pyrimidique des cytosines. Il s'agit de DNMT1, DNMT3A et DNMT3B. Ces 3 enzymes s'organisent en deux régions fonctionnellement distinctes. La région régulatrice amino-terminale, de longueur variable, contient de nombreux motifs permettant l'interaction des DNMTs avec une multitude de molécules chromatinienne ainsi que leur ciblage à différents endroits du génome. Le domaine PWWP de DNMT3A et 3B leur permet par exemple d'être recrutées dans l'hétérochromatine constitutive, au niveau des régions péricentromériques. Le domaine catalytique carboxy-terminal contient quant à lui un motif méthyltransférase conservé entre les 3 enzymes.

La distribution différentielle de DNMT1, DNMT3A et DNMT3B à travers le génome témoigne des fonctions spécifiques à chacune de ces enzymes et d'une redondance uniquement partielle.

DNMT1 est essentiellement retrouvée au niveau de la fourche de réplication (166, 167). Elle y est recrutée via son interaction avec PCNA (Proliferating Cell Nuclear Antigen) et UHRF1 (Ubiquitin-like, containing PHD and RING finger domains, 1) qui possède un domaine de liaison SRA (SET and RING Associated) à l'ADN hémiméthylé. DNMT1 recopie le méthylome sur les cytosines du brin d'ADN néo-synthétisé lors de la réplication cellulaire, elle est ainsi qualifiée de méthyl-transférase de maintenance (**Figure 15**).

A l'inverse, DNMT3A et DNMT3B transfèrent un groupement méthyl sur des résidus non méthylés, et sont donc qualifiées d'ADN méthyl-transférases *de novo*. Elles ont des fonctions partiellement non redondantes puisqu'elles sont par exemple toutes les deux essentielles au développement embryonnaire. En effet, les KO *Dnmt3a*^{-/-} et *Dnmt3b*^{-/-} sont individuellement létaux et l'arrêt de l'embryogénèse ne se fait pas au même stade dans les deux mutants (168). Les mécanismes de ciblage des DNMT *de novo* peuvent dépendre, ou non, de la séquence d'ADN (**Figure 15**). Les DNMT3 peuvent être recrutées au niveau de l'hétérochromatine grâce à leur domaine PWWP par un mécanisme ne dépendant pas de la reconnaissance d'une séquence consensus d'ADN. Le domaine PWWP reconnaît en effet l'histone H3 triméthylée sur sa lysine 36 (H3K36me3) (169), une modification épigénétique retrouvée au niveau du corps de gène transcriptionnellement actifs mais aussi plus récemment dans des régions fermées du génome (170). Cette distribution de la marque est en accord avec les méthylomes décrits dans

la plupart des études. Les DNMT *de novo* peuvent également être recrutées de façon spécifique de la séquence d'ADN, soit par l'intermédiaire de facteurs répresseurs de la transcription, comme c'est le cas avec cMyc qui recrute DNMT3A au niveau du promoteur du gène *p21^{cip1}*, soit par l'intermédiaire de la machinerie d'interférence à l'ARN. Ce dernier mécanisme identifié chez les plantes reste flou chez les mammifères. Néanmoins, la délétion de *Dicer*, un composant important de la machinerie d'interférence à l'ARN, se traduit par l'expression aberrante d'éléments répétés localisés dans les régions péri-centromériques, et par une baisse globale de la méthylation de l'ADN.

Si les DNMT1, DNMT3A et DNMT3B ont des localisations différentes et des fonctions partiellement non redondantes, elles sont toutes trois nécessaires pour générer un méthylome complet lors de l'embryogenèse et on sait maintenant qu'elles peuvent coopérer. En effet il a été montré *in vitro* que la méthylation par DNMT3A stimulait la méthylation *de novo* par DNMT1 (171). Une étude comparative du méthylome de cellules souches embryonnaires de type sauvage ou déficientes pour *Dnmt1*, ou *Dnmt3a* ou *Dnmt3b* a permis d'attribuer à chacune des enzymes sa part respective dans la méthylation *de novo* de l'ADN (172). Enfin, les DNMT3 participent également à la maintenance du méthylome lors de la réplication cellulaire. Sur certaines régions du génome, comme le promoteur des séquences LINE, DNMT1 est en effet peu efficace pour maintenir la méthylation et DNMT3A/B semble nécessaire pour transférer un groupement méthyl *de novo* sur ces régions (173).

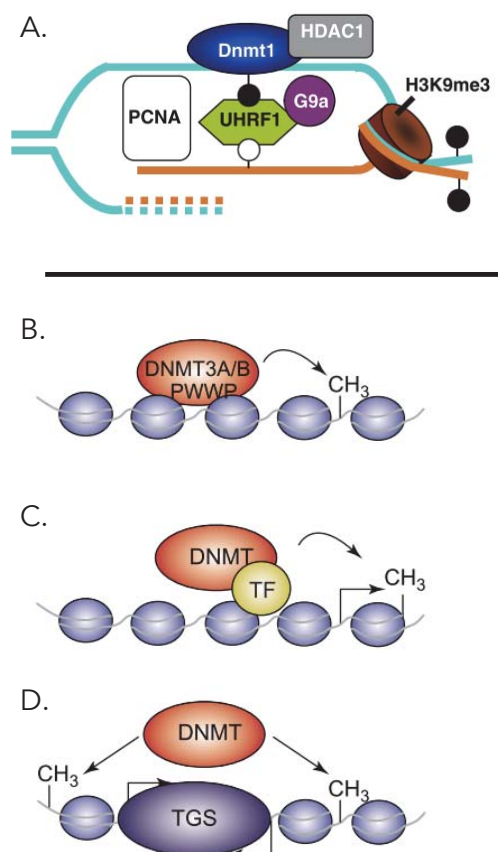


Figure 15 : Mécanismes de ciblage des DNMT à l'ADN

A. DNMT1 est recrutée au niveau de la fourche de réplication par UHRF1 et PCNA afin de reproduire le patron de méthylation sur le brin néo-synthétisé.

B-D. Les DNMT *de novo* peuvent être recrutées selon divers mécanismes. Leur domaine PWWP peut reconnaître des modifications post-traductionnelles de la chromatine et cibler les DNMT3 dans l'hétérochromatine (B). Les DNMT peuvent aussi être ciblées par l'intermédiaire d'un facteur de transcription (C) ou par la machinerie d'interférence à l'ARN (D).

Adapté de (174, 175)

2. Rôle de la méthylation de l'ADN

Comme énoncé précédemment, la fonction la plus décrite de la méthylation de l'ADN est la répression de la transcription. Elle impacte des processus biologiques variés, parfois interdépendants, allant de la différenciation cellulaire à la répression des éléments transposables. Cette répression implique différents mécanismes reposant soit directement sur la présence du groupement méthyl qui interfère avec la liaison de facteur de transcription, soit nécessitant un « chromatin reader » comme intermédiaire de répression (**Figure 16**).

Tout d'abord, les facteurs *trans* peuvent être sensibles à la méthylation des cytosines au niveau de leur séquence de liaison. C'est par exemple le cas du répresseur transcriptionnel CTCF (CCCTC-binding factor). Il est impliqué dans la régulation de l'expression des gènes adjacents, *Igf2* et *H19*, placés sous le contrôle d'un même enhancer soumis à l'empreinte parentale. L'enhancer est non méthylé sur l'allèle maternel ce qui permet à CTCF de s'y fixer et de permettre ainsi l'expression mono-allélique du gène *Igf2*. A l'inverse, sur l'allèle paternel, l'enhancer est méthylé et CTCF ne peut s'y fixer ce qui aboutit à l'expression d'*H19* mais pas d'*Igf2* (176).

La liaison des 5mC par des « chromatin reader » peut permettre la répression transcriptionnelle en interférant avec la fixation d'un facteur de transcription, en masquant sa séquence de liaison ou en recrutant des complexes répresseurs de la chromatine par exemple. Parmi les « chromatin reader » des 5mC les plus étudiées, on retrouve les protéines MeCP2, MBD1, MBD2 et MBD4 qui partagent un domaine (MBD) conservé entre les espèces permettant leur liaison au 5mC des di-nucléotides CpG (177). Les protéines UHRF1 et UHRF2 peuvent reconnaître les 5mC via un autre domaine conservé : le domaine SRA (SET and RING Associated) (178). Enfin les 5mC sont également reconnues de façon dépendante de la séquence d'ADN par des facteurs de transcription tels KAISO et KLF4 via leur domaine en doigts de zinc respectifs (179, 180). Récemment, des cartographies à grande échelle des protéines liées aux 5mC par spectrométrie de masse ou microarray protéique ont permis de mettre en évidence un grand nombre de MeCP dont la mobilisation est dépendante du type cellulaire (178, 180, 181). De plus les DNMT, outre leur fonction enzymatique, peuvent également se comporter comme des « chromatin reader ». Elles permettent notamment le recrutement d'autres facteurs impliqués dans le remodelage de la chromatine (182). Par exemple MeCP2 permet le recrutement d'histone déacétylases qui enlèvent les groupements permissifs acétyles de la chromatine. Néanmoins le déroulement de ce remodelage induit par les partenaires des MeCP reste flou (183).

Il est intéressant de noter que la méthylation sur le corps des gènes est corrélée à l'absence de méthylation sur leur promoteur et à leur état transcriptionnellement actif. A l'inverse quand le promoteur est inactivé par méthylation le corps du gène est lui déméthylé. Ce pattern opposé de méthylation a soulevé l'hypothèse que la méthylation des cytosines intra-géniques pourrait permettre de réprimer l'initiation d'une transcription d'éléments transposables chevauchant un gène exprimé (184, 185). Plus récemment, il a été suggéré que la méthylation de l'ADN sur le corps des gènes permet d'éviter la liaison et l'initiation de la transcription inappropriée de l'ARN polymérase II sur le corps du gène. Néanmoins la présence d'éléments transposables intragéniques n'a pas été investiguée par les auteurs (186, 187). Shukla et collaborateurs ont d'autre part proposé que la méthylation du corps des gènes permettrait l'inclusion d'exons alternatifs (188). Cependant cette étude ne porte que sur un seul gène. Le rôle de la méthylation de l'ADN sur le corps du gène reste donc à approfondir.

A.

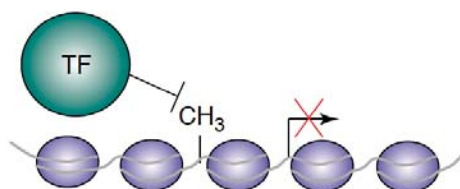
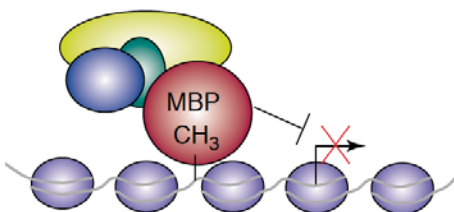


Figure 16 : Mécanismes répresseurs de la transcription par la méthylation de l'ADN

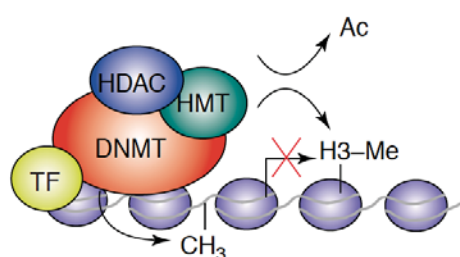
A. La méthylation de l'ADN interfère avec la liaison d'un facteur de transcription (TF).

B.



B. La liaison de « chromatin readers » au 5mC interfère avec la liaison de complexe trans-activateur.

C.



C. Les « chromatin readers » recrutent des protéines impliquées dans le remodelage. La compaction de la chromatine ne permet pas la transcription.

Adapté de (174)

3. La méthylation de l'ADN est réversible

La méthylation de l'ADN est spécifique du type cellulaire. Les sites méthylés dans les cellules souches embryonnaires ne le sont pas forcément dans les cellules différenciées et inversement (183). La méthylation des cytosines est donc réversible. Elle peut disparaître par dilution au cours des divisions cellulaires ou être activement enlevée par des enzymes (**Figure**

17). En effet les enzymes de la famille TET peuvent oxyder de façon itérative les 5mC pour générer des 5-hydroxyméthylcytosines (5hmC), puis des 5-formylcytosines (5fC) et enfin des 5-carboxycytosines (5caC) (189–191). Le modèle exact de déméthylation active des cytosines n'est pas connu avec certitude. Il semblerait que les 5caC puissent être excisées par l'ADN glycosylase TDG puis remplacées par des cytosines grâce au complexe BER (Base Exision Repair). Les intermédiaires d'oxydation peuvent recruter des « chromatin readers » distincts (dont des régulateurs transcriptionnels). Ainsi, les intermédiaires de dégradation des 5mC, en plus de leur fonction essentielle dans la dynamique du méthylome, pourraient également avoir un rôle actif dans la fonction du génome à l'instar d'autres marques épigénétiques (178).

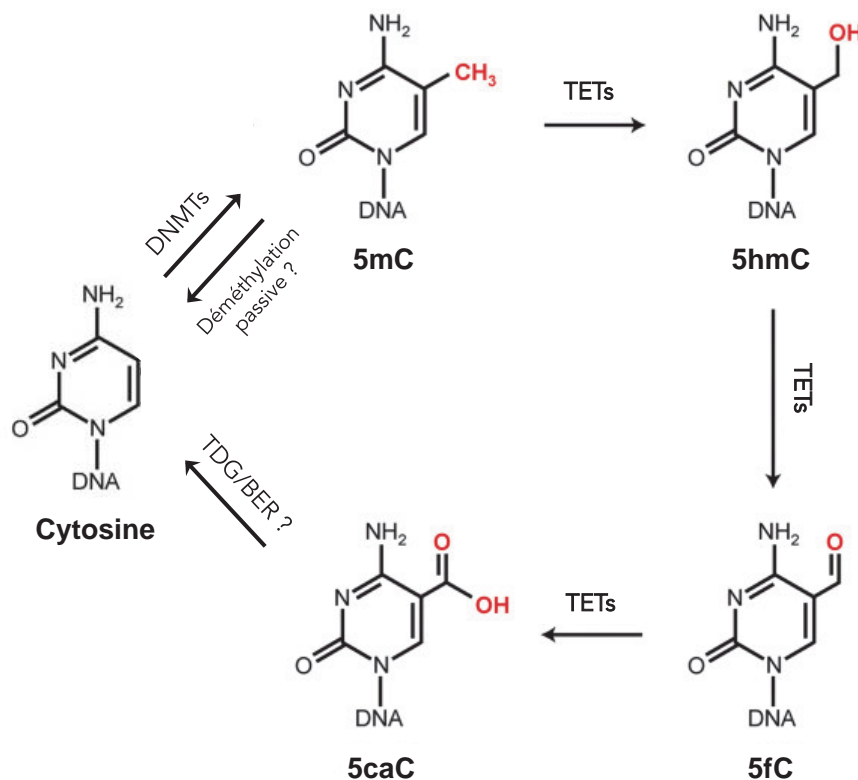


Figure 17 : Méthylation des cytosines

Les ADN méthyl transférases (DNMTs) permettent le transfert d'un groupement méthyl sur le carbone 5 des cytosines (5mC). La méthylation des cytosines est un processus réversible de façon passive ou active grâce aux enzymes de la famille TET. Elles peuvent hydrolyser le groupement méthyl en groupement hydroxy-méthyl puis formyle et enfin en groupement carboxyle pour générer respectivement des 5hmC, 5fC et 5caC. Ces dernières pourraient être excisées par l'ADN glycosylase TDG et remplacées par des cytosines non modifiées grâce au complexe de réparation de l'ADN BER.

C. Modification post-traductionnelles des histones

1. Codes histones

Les histones sont des protéines basiques, riches en résidus lysine et arginine. Elles sont la cible de modifications post-traductionnelles covalentes (MPT) qui sont essentiellement focalisées sur leur queue amino-terminale, même si leur domaine globulaire ou leur région carboxy-terminale peuvent aussi être ciblées (192) (**Figure 18**). Les modifications qui affectent les quatre histones de cœur, notamment H3 et H4, ont un rôle clef dans la régulation des fonctions du génome. En servant de point d'ancrage à de nombreux « chromatin readers », tels que les protéines HP1 qui se lient à H3K9me3, elles contrôlent par exemple qualitativement et quantitativement la transcription génique. L'étude de leur effet combiné a permis de déchiffrer un « code histone » qui permet de déterminer la partition jouée par le génome (193, 194).

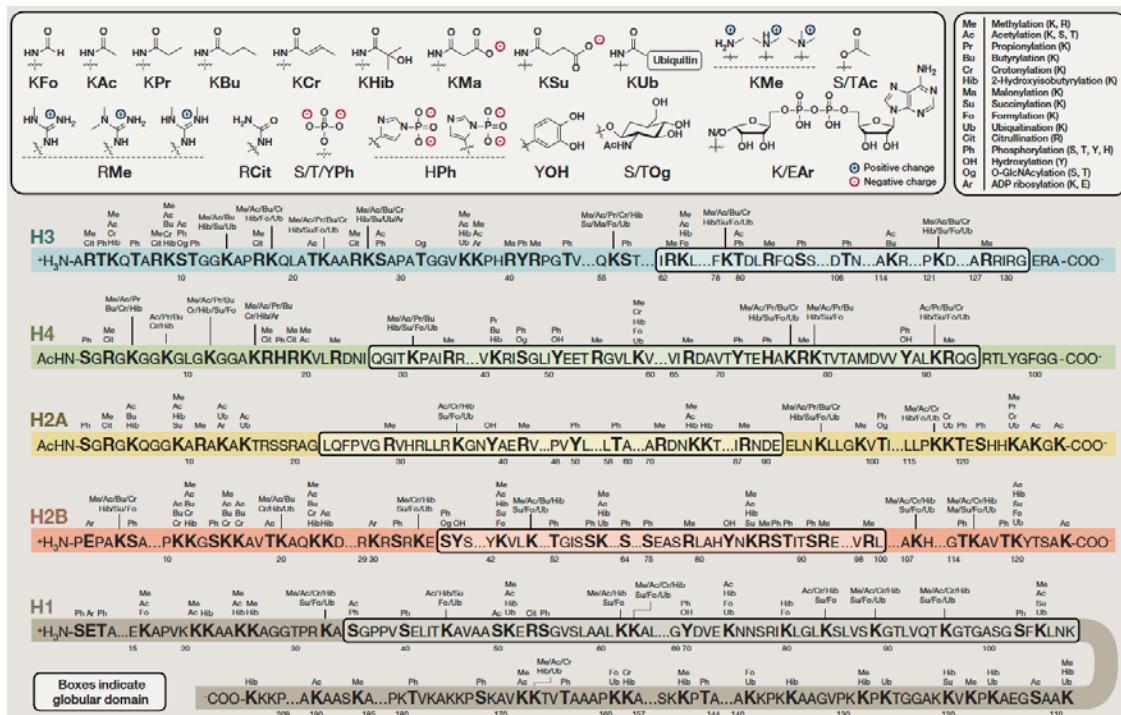


Figure 18 : Modifications post-traductionnelles des histones

Des modifications post-traductionnelles peuvent être déposées sur des résidus particuliers, essentiellement Lysine (K), Sérine (S), Arginine (R) ou Thréonine (T), des queues amino- et carboxy-terminales des histones. Leur domaine globulaire, représenté ici par un rectangle noir, peut aussi être ciblé par les « chromatin writers », mais les effets biologiques associés à ces modifications sont encore peu caractérisés. L'acétylation, la phosphorylation, la méthylation et l'ubiquitination font partie des marques les plus étudiées. (192)

Si certaines marques peuvent être simultanément présentes sur une même région du génome, c'est par exemple le cas d'H3K4me1 et d'H3K27ac sur les enhancers actifs, d'autres sont mutuellement exclusives car ciblant un même résidu comme H3K9me3 et H3K9ac. La présence d'une MPT peut aussi impacter la déposition d'une autre marque ou altérer le recrutement des « chromatin readers » et donc la fonctionnalité de la marque. C'est le cas avec la phosphorylation de la sérine 10 de l'histone H3 (H3S10P) qui interfère avec la liaison des protéines HP1 sur H3K9me3. La déposition de cette marque est aussi un prérequis pour que l'histone acétyltransférase GCN5 puisse acétyler H3K14 et ainsi initier l'assemblage du complexe d'initiation de la transcription (195).

La combinatoire des MPT d'histone permet donc d'identifier les éléments fonctionnels du génome et de déterminer leur statut d'activation. Par exemple, les enhancers peuvent être identifiés en cartographiant les régions enrichies en H3K4me1 (196). Le statut actif de l'enhancer peut être déterminé par la présence supplémentaire d'H3K27ac, ou de l'histone acétyltransférase p300 qui la catalyse, sur les régions ayant été identifiées comme enrichies en H3K4me1 (197). A l'inverse un enhancer également enrichi en H3K27me3 sera en pause (« poised »), prêt à être activé ou réprimé selon les informations intégrées par la cellule.

J'aborderai brièvement l'acétylation et la méthylation des histones avant de focaliser mon propos sur la méthylation d'H3K9.

2. Acétylation des histones

De nombreux résidus peuvent être acétylés par des Histones AcétylTransférases (HAT). Outre les lysines, il a en effet été montré que ces enzymes peuvent transférer le groupement acétyl de l'acétyl-coenzyme A sur d'autres acides aminés tels que les sérines ou les thréonines (198). Les HAT de type I sont localisées dans le cytoplasme. Elles acétylent notamment les histones nouvellement synthétisées qui sont ensuite transportées dans le noyau et incorporées à la chromatine. Les HAT de type II sont nucléaires et acétylent les histones chromatinienne. La spécificité de résidus cibles des HAT peut être restreinte ou multiple. Elle dépend entre autres des protéines avec lesquelles elles interagissent. L'acétylation des lysines neutralise la charge positive de l'acide aminé et diminue ainsi la force d'interaction entre l'ADN, chargé négativement, et le nucléosome. Il s'agit donc d'une marque exclusivement permissive, associée à la chromatine décondensée, ou euchromatine, et donc à des loci transcriptionnellement actifs.

L'acétylation est une marque qui peut aussi être activement enlevée par des histones déacétylases (HDAC). Moduler l'expression des HAT ou des HDAC aboutit systématiquement à une perturbation de la distribution de la marque, ce qui suggère que sa présence ou son absence dépend d'un équilibre subtil et dynamique entre ces deux types d'enzyme. Chez l'Homme, 18 HDAC distinctes ont été identifiées à ce jour. Elles sont réparties en quatre familles. Les HDAC de classe I et II dé-acétylent les résidus histones via l'utilisation d'un ion zinc. Les HDAC de classe III utilisent quant à elles le cofacteur NAD⁺ pour enlever le groupement acétyle. L'HDAC 11, la seule HDAC de classe IV, partage des homologies de séquence avec les HDAC de classe I et II, on peut donc émettre l'hypothèse que le mécanisme catalytique nécessite un ion zinc (199).

L'acétylation des lysines, si elle a un rôle permissif de par la neutralisation de charge qu'elle provoque, peut également être « lue » par des « chromatin readers » possédant un domaine « Bromo ». Ces protéines permettent ensuite le recrutement d'autres facteurs permettant la transcription. Une cascade moléculaire décrite par Agalioti et collaborateurs illustre bien le rôle activateur de l'acétylation des lysines sur l'expression génique. Ces auteurs se sont intéressés à la réponse IFN de type I déclenchée suite à une infection virale. Ils ont plus précisément essayé de déterminer comment le gène *Ifnb* était induit dans des cellules infectées par le Virus Sendai. Il avait déjà été montré que ce gène est sous le contrôle d'un enhancer situé 110 à 45 pb en amont du TSS. Suite à une infection, NF-κB, des membres de la famille des IRFs, et l'hétérodimère ATF-2/c-Jun coordonnent sur cet élément génomique le recrutement d'un complexe multimérique, appelé enhanceosome, nécessaire à la transcription à haut niveau du gène. Agalioti et collaborateurs ont montré que l'HAT GCN5 faisait partie des modificateurs chromatinien recrutés au niveau de ce complexe. Cette dernière acétyle les résidus H3K9 au niveau du TSS alors qu'une kinase, non identifiée dans l'article, phosphoryle H3S10, permettant ainsi à GCN5 d'acétyle à son tour la lysine 14 de l'histone 3. Ces trois marques permettent ensuite le recrutement de « chromatin readers » à domaine Bromo, notamment le facteur d'initiation de la transcription TFIID et le complexe de remodelage de la chromatine SWI/SNF, tous deux nécessaires au déplacement du nucléosome et à l'assemblage de la machinerie transcriptionnelle au niveau du TSS (195).

L'acétylation des résidus H3K27 et H3K9 sont des marques épigénétiques très utilisées dans les études de génomique pour identifier les régions actives de la chromatine, telles que les promoteurs des gènes transcrits et les enhancers actifs.

3. Méthylation des histones

La méthylation est une marque épigénétique plus complexe dont l'impact biologique dépend de l'histone et du résidu ciblé, et du nombre de groupements méthyl déposés. Si elle peut être catalysée sur différents résidus, les lysines telles que H3K4, H3K9, H3K27, H3K36 et H4K20 constituent néanmoins des cibles préférentielles. Lorsqu'elle est retrouvée sur H3K4 et H3K36, la méthylation est globalement associée à une chromatine ouverte et donc à une activité transcriptionnelle intense. Il faut cependant noter que cette « règle » n'est pas absolue puisque des études par ChIP-seq ont montré que H3K36me3 est aussi enrichie dans des zones d'hétérochromatine facultative et constitutive. Lorsqu'elle est déposée sur H3K27, H4K20 ou H3K9, la tri-méthylation est, elle, globalement associée à l'hétérochromatine. A nouveau, des exceptions existent puisque H3K9me3 a été détectée sur le corps de gène transcriptionnellement actif où elle pourrait avoir un impact qualitatif et/ou quantitatif sur la transcription (voir ci-dessous). Selon le résidu ciblé, mais aussi l'élément génomique sur lequel elle est déposée, la méthylation peut donc avoir des effets différents, parfois antagonistes ! Les résidus peuvent de plus être marqués par un à trois groupements méthyl (me1, me2 ou me3) (**Figure 19**). Ceci complexifie encore le « code histone » puisque la mono- ou la tri-méthylation peuvent parfois avoir des localisations ou des effets biologiques différents. Ainsi, les enhancers sont par exemple majoritairement enrichis en H3K4me1 alors que les promoteurs le sont essentiellement en H3K4me3 (196).

Les « chromatin writers » impliqués dans la méthylation des histones sont essentiellement des histones méthyl transférases (HMT). Parmi elles, les lysine méthyl transférases (KMT) sont spécialisées dans la méthylation des lysines. Quel que soit le résidu qu'elles ciblent, elles ont toutes en commun un domaine SET (**S**uppressor of variegation 3-9 (Su(var)3-9), **E**nhancer of zeste (E(z)) and **T**rithorax (Tx)) qui catalyse le transfert d'un groupement méthyl à partir du cofacteur S-Adenosyl-L-Méthionine (SAM) sur une lysine. Ce domaine est encadré par les domaines pré-SET et post-SET nécessaires à l'activité catalytique. En plus d'un rôle structurel, ils sont notamment impliqués dans l'interaction de l'enzyme avec les queues des histones et le cofacteur SAM. Les KMT possèdent enfin au moins un autre domaine leur permettant de cibler la chromatine directement ou via l'interaction de partenaires moléculaires tels que HP1.

La méthylation des histones, comme l'acétylation, est une marque dynamique qui peut être retirée par des enzymes spécifiques, les histones déméthylases (HDM). Elles contiennent un domaine catalytique « Jumonji C » (JmjC) leur permettant de déméthyle les histones. Il

existe plus d'une dizaine de KDM identifiées à ce jour avec des spécificités multiples pour les résidus histone et aux fonctions partiellement redondantes.

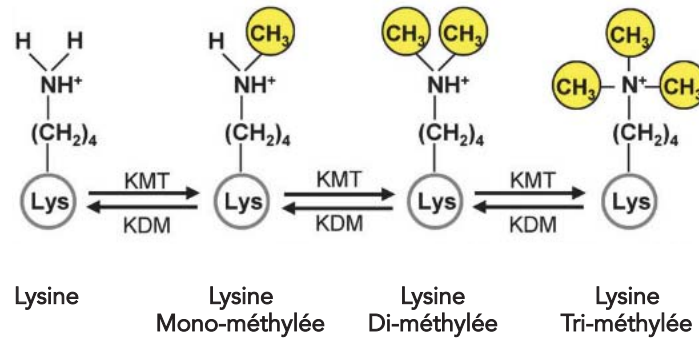


Figure 19 : Méthylation des lysines

Les résidus lysine peuvent être méthylés par des lysine méthyl transférase (KMT) selon un processus itératif. Des lysine déméthylases (KDM) peuvent enlever les groupement méthyls. Adapté de (200).

a. Méthylation d'H3K4

H3K4 peut être mono-, di-, ou tri-méthylée. Les principales KMT responsables de la méthylation d'H3K4 appartiennent aux familles protéiques « Mixed Lineage Leukemia », « Suppressor of variegation enhancer of zeste, Trithorax » et « SET et MYND domaine ». Elles ont des fonctions partiellement redondantes et sont souvent incluses dans des complexes macromoléculaires. C'est le cas de MLL1 et MLL2 qui font partie des complexes activateurs Trithorax.

La méthylation d'H3K4 est essentiellement associée à un état permissif de la chromatine. H3K4me3 est enrichie au niveau du TSS des gènes (196). Cette marque peut également être retrouvée au niveau d'enhancer mais s'y accumule moins. A l'inverse H3K4me1 est essentiellement enrichie au niveau des enhanceurs et le signal obtenu par ChIP-seq est corrélé avec le niveau d'activation de l'élément cis-régulateur au sein de la population cellulaire étudiée (196). Lors d'analyse épigénétique, ces deux marques sont souvent combinées afin de caractériser au mieux les éléments génomiques. Ainsi, un ratio H3K4me1/H3K4me3 élevé est caractéristique d'enhancers alors qu'un ratio faible est plutôt retrouvé au niveau des régions promotrices des gènes. H3K4me1/me2/me3 sont également retrouvées dans une moindre mesure sur le corps des gènes exons et introns mais cette distribution n'a pas été associée à

une fonction particulière. Enfin des études rapportent un enrichissement sur des *loci* qui accumulent également des marques essentiellement répressives telles que H3K27me3 et H3K9me3 (201–203). Dans ce cas, la présence commune de ces marques ambivalentes déterminerait des *loci* dans un état chromatinien « poised » ni actifs, ni complètement verrouillés.

De nombreuses KDM ont été impliquées dans la déméthylation d'H3K4. La première à avoir été identifiée, KDM1 (ou LSD1) est impliquée dans la déméthylation d'H3K4me1/me2 mais peut également cibler H4K20me1/me2 et H3K9me1/me2. Son action portant sur trois marques épigénétiques aux rôles distincts complexifie l'étude de ses fonctions. Cependant des études suggèrent une implication dans la balance entre renouvellement cellulaire et différenciation.

b. Méthylation d'H3K27

H3K27me3 est essentiellement associée à de la répression génique. Elle est générée par les sous-unités catalytiques EZH1 et EZH2 du complexe répresseur Polycomb 2 (PRC2) qui seul n'a pas d'activité enzymatique. Ce complexe comprend également les sous-unités SUZ1/2, EED et Rbap46/48 et peut interagir avec d'autres protéines accessoires. SUZ1/2 et EED sont nécessaires pour que PRC2 puisse se lier à H3K27me3 et ainsi propager la marque. H3K27me3 est retrouvée au niveau des promoteurs et des enhanceurs. Seule, elle est associée à leur répression. Si elle co-localise avec H3K4me1/me3, H3K27me3 maintient alors ces régions cis-régulatrices dans un état « poised ». H3K27me3 participe également à l'inactivation du chromosome X.

Plusieurs KDM ont été impliquées dans la déméthylation d'H3K27me3. L'une d'entre elles, KDM6A (ou UTX), fait partie d'un complexe multimérique comprenant la lysine méthyl transférase MLL4 suggérant ainsi que la déméthylation d'H3K27 et la méthylation d'H3K4 sont associées au sein d'un complexe activateur de la chromatine.

Si H3K27me2 a la même distribution qu'H3K27me3 et est essentiellement associée à la répression génique, H3K27me1 est enrichie en aval des promoteurs de gènes actifs. Cette distribution suggère qu'elle aurait un rôle activateur de l'expression génique (204).

D. Méthylation d'H3K9

1. Mono-, di- ou tri-méthylation d'H3K9

H3K9 peut être mono-, di- et tri-méthylée. H3K9me2 et H3K9me3 ont des distributions relativement similaires et sont essentiellement associées à une conformation réprimée de la chromatine et à de la répression génique. En revanche H3K9me1, comme H3K27me1, est enrichie au niveau des TSS de gènes exprimés (204). Elle s'accumule aussi sur les séquences insulatrices (204). Le lien causal entre la déposition d'H3K9me1 sur ces régions génomiques et leur fonction n'a pas encore été établie. Par la suite de ce manuscrit, je me focaliserai donc sur la tri-méthylation d'H3K9.

2. Localisation et fonctions d'H3K9me3

Les fonctions d'H3K9me3 sont multiples et dépendent de sa localisation dans le génome. Historiquement, cette marque a été décrite comme ayant un rôle structurel et fonctionnel majeur dans l'hétérochromatine constitutive (205, 206). Elle a depuis été identifiée dans l'euchromatine et l'hétérochromatine facultative où elle semble avoir des effets parfois antagonistes, notamment sur la transcription. Lorsqu'elle se dépose sur les régions promotrices ou enhancers, elle est impliquée dans la répression de la transcription alors que son accumulation sur le corps des gènes pourrait avoir un impact positif ou qualitatif sur leur expression. H3K9me3 joue aussi un rôle clef dans la répression des éléments transposables et dans la réparation de l'ADN (207, 208).

a. Composant de l'hétérochromatine

H3K9me3 est enrichie dans l'hétérochromatine constitutive, notamment péricentromérique chez la levure (209) et les mammifères (206). Les souris déficientes pour les enzymes SUV39H1 et SUV39H2, qui jouent un rôle clef dans sa déposition dans ces régions, sont peu viables. L'absence d'H3K9me3 dans les régions péricentromériques provoque en effet une instabilité chromosomique qui est à terme létale. Les régions répétées sub-télomériques et télomériques sont également composées d'hétérochromatine et sont donc

également enrichies en H3K9me3. La perte de la marque est associée à une élongation des chromosomes augmentant le risque de recombinaison homologue (210).

b. Répression des promoteurs

H3K9me3 est également retrouvée au niveau de l'hétérochromatine facultative et de l'euchromatine où elle est associée à la cis-régulation de l'expression des gènes. En s'accumulant sur les promoteurs, H3K9me3 réprime la transcription. Sa déposition induit le recrutement de la protéine HP1 α (Heterochromatin Protein 1) qui initie localement la formation de l'hétérochromatine. Cela a notamment été montré dans des lignées cellulaires exprimant des gènes rapporteurs et dans des cellules murines primaires (211, 212).

La déposition d'H3K9me3 sur le promoteur du gène bivalent maintient l'ARN polymérase II en « stand-by ». Par exemple, dans les cellules souches embryonnaires trophoblastiques et mésenchymateuses (ou pré-adipocytes), des domaines bivalents pour H3K4me3 et H3K9me3 associés à de la répression transcriptionnelle ont été décrits (201, 202). Ces domaines bivalents existent également dans les cellules différenciées, mais leur distribution dépend alors du type cellulaire considéré et de son niveau de maturation, différenciation, activation... (202). Ces domaines « poised » permettent de préparer les cellules à exprimer rapidement un groupe de gènes ou à l'inverse à les réprimer en fonction des signaux environnementaux. En effet, Matsumura et collaborateurs ont montré dans les pré-adipocytes que l'accumulation d'H3K9me3 en aval des TSS des gènes *Cebpa* et *Pparg*, impliqués dans l'adipogénèse, corrèle avec l'accumulation d'une forme non-processive de l'ARN polymérase II sur leur promoteur. En revanche, lorsque les pré-adipocytes perçoivent un stimulus de différenciation, H3K9me3 n'est plus enrichie sur ces gènes et l'ARN polymérase II s'accumule sur le corps du gène. La déplétion d'H3K9me3 par traitement des pré-adipocytes avec un ARN anti-sens dirigé contre SETDB1, l'une des enzymes responsables de la déposition de la marque, induit la différenciation spontanée des pré-adipocytes en adipocytes.

La déposition d'H3K9me3 sur les régions promotrices participe aussi à la détermination d'une cellule dans un lignage donné. Il a par exemple été montré dans les précurseurs des cellules souches hématopoïétiques (HSPC), qu'H3K9me3 s'accumule sur des gènes impliqués dans la néoglucogénèse. Cette fonction métabolique est restreinte au foie, aux reins et aux muscles (213). En absence d'H3K9me3 sur leurs promoteurs, ces gènes sont exprimés et déstabilisent ainsi la fonction cellulaire. De plus, la modulation dynamique d'H3K9me3 est impliquée dans la déprogrammation des fibroblastes embryonnaires murins (MEF) pour induire

des cellules souches pluripotentes (iPSC) (214). Dans les pre-iPSC, H3K9me3 est enrichie sur les gènes de pluripotence et les réprime. Il faut l'intervention de lysine déméthylases qui enlèvent la marque pour que les pre-iPSC se déprogramment complètement et deviennent des iPSC.

Ainsi, la déposition d'H3K9me3 sur les régions promotrices est essentielle au développement et à la différenciation cellulaire en réprimant des gènes inductibles, mais elle est également primordiale pour réprimer les gènes associés à une programmation inappropriée des cellules. Cette marque constitue donc une barrière épigénétique nécessaire au maintien de l'identité et de l'intégrité cellulaire.

c. Répression des enhancers

H3K9me3 peut aussi être déposée sur les enhancers. Selon les autres marques présentes au niveau de ces éléments génomiques, leur statut oscille alors entre « réprimé » et « poised » (215). Le groupe du Dr. Sacconi a par exemple mis en évidence qu'H3K9me3 réprime les enhancers de façon spécifique du type cellulaire (216). Les auteurs se sont appuyés sur l'expression différentielle des gènes *Mdc* et *Il12b* entre deux populations cellulaires. Ces deux loci sont en effet induits par NF- κ B dans les cellules dendritiques après stimulation au LPS mais pas dans les fibroblastes 3T3 après stimulation au TNF- α , et ce, même si cette cytokine pro-inflammatoire active aussi l'hétérodimère p50/p65. Dans les fibroblastes, l'absence d'expression des gènes s'accompagne d'une accumulation d'H3K9me3 sur les régions qui entourent les enhancers, ce qui n'est pas le cas dans les cellules dendritiques. En revanche, l'expression de la lysine déméthylase Jumanji D2d (*Jmjd2d*) est induite dans les cellules dendritiques et son recrutement sur les régions flanquant les enhancers inhibe l'accumulation de la marque. Ceci suggère que la distribution d'H3K9me3 sur les enhancers situés dans l'euchromatine et l'hétérochromatine facultative est spécifique à chaque type cellulaire et résulte d'un équilibre subtil entre mécanismes de déposition et d'élimination. En effet la délétion de *Jmjd2d* dans les cellules dendritiques se traduit par une accumulation d'H3K9me3 sur ces régions et par la répression de l'expression de *Mdc* et d'*Il12b*. A l'inverse, l'expression ectopique de *Jmjd2d* dans les fibroblastes induit une perte d'H3K9me3 et l'expression des gènes normalement restreints aux cellules dendritiques. Cette étude révèle donc un rôle causal et non redondant pour H3K9me3 dans la répression d'enhancers contrôlant l'expression de gènes inductibles. De plus, en généralisant leur donnée à l'échelle du génome, les auteurs ont mis en évidence que les enhancers réprimés par H3K9me3 étaient principalement ciblés par des facteurs de transcription « ubiquitaire ». Cette marque permet donc de définir, dans chaque cellule et parmi l'ensemble du répertoire qu'ils peuvent cibler, le réseau d'enhancers

sur lesquels ces facteurs de transcription peuvent se déposer. Le mécanisme de ciblage d'H3K9me3 permettant la répression des enhancers de façon type cellulaire reste cependant à identifier.

d. Répression des éléments transposables

H3K9me3 est impliquée dans la répression des éléments transposables (TE) dans différents types cellulaires. Les TE représentent près de 40% des génomes murins et humains (217). Ils proviennent d'invasions successives du génome, principalement rétrovirales, ayant eu lieu au cours de l'évolution. Si certains de ces éléments sont dégénérés et ne peuvent plus se transposer, d'autres plus récemment intégrés au génome ont maintenu cette capacité. On peut globalement subdiviser les rétroéléments en trois catégories. Les TE à LTR (Long Terminal Repeat) regroupent essentiellement trois classes de rétrovirus endogènes (ERV) et comptent pour près de 10 % des génomes murins et humains. Les régions LTR qui flanquent le génome viral ont une activité cis-régulatrice. Les TE sans LTR, les LINE (Long-Interspersed Nucleotide Element) et les SINE (Short-Interspersed Nucleotide Element), comptent pour près de 30% du génome. Enfin, les transposons à ADN se déplacent dans le génome grâce à l'expression d'une transposase leur permettant d'être coupés et collés ailleurs dans le génome. La réplication des ERV, des LINE et des SINE passe par un intermédiaire d'ARN. Dans les cellules souches embryonnaires murines (mESC) H3K9me3 est enrichie sur les ERVs de classe II et III. Il a été proposé que la répression dépendante d'H3K9me3 cible des ERV différents de ceux réprimés par la méthylation de l'ADN (218). Néanmoins, sur certains éléments, les deux mécanismes pourraient agir de façon synergique (219). L'accumulation d'H3K9me3 est notamment nécessaire à la répression d'ERVs de classe II actifs appelés IAP (intracisternal A-type particule). Il a en effet été montré qu'en absence du corépresseur transcriptionnel TRIM28 ou de la lysine méthyl transférase SETDB1 le niveau d'H3K9me3 diminue sur ces régions et les IAP sont exprimés (207, 218, 220, 221). Dans les mESC, ceci initie l'expression anormale de transcrits chimériques à partir des LTR lorsque les ERVs sont situés à proximité de gènes (218). Il a été suggéré que les LTR pourraient ainsi être des éléments régulateurs de l'expression génique. Cette hypothèse est en accord avec d'autres études. Dans l'une d'elles, les auteurs ont tiré parti de deux lignées cellulaires de mESC provenant de souris de fonds génétiques différents et n'ayant donc pas les mêmes IAP pour montrer que la marque H3K9me3 réprimant les IAP peut être propagée sur les gènes adjacents et les réprimer (222). Le groupe du Pr. Trono, en utilisant des mESC déficientes pour TRIM28, a ensuite montré que ce mécanisme existait à l'échelle du génome dans des mESC. La perte d'H3K9me3 sur les ERV en absence de TRIM28 favorise l'accumulation d'H3K4me1 et d'H3K27ac, permettant ainsi l'expression

génique (223). Enfin, H3K9me3 permet également la répression des éléments LINE dans les mESC (224).

Le groupe du Pr. Jenuwein a suggéré que la répression des ERVs par H3K9me3 était un mécanisme spécifique aux mESC car le signal d'H3K9me3 y est plus élevé que dans des cellules neurales progénitrices (NPC) ou dans des fibroblastes embryonnaires murins (iMEF). Cependant, des études récentes ont démontré l'importance de ce mécanisme dans des cellules différenciées. En effet dans les lymphocytes B et pro-B, H3K9me3 est nécessaire pour réprimer les familles d'ERV MLV, MMTV et VL30 (225, 226). De façon remarquable, les familles d'ERVs ciblées par H3K9me3 dans les LB ne sont pas les mêmes que dans les mESC. Cette spécificité renforce ainsi l'hypothèse que les LTR agiraient comme des éléments régulateurs de l'expression des gènes de façon spécifique du tissu ou du type cellulaire. Enfin, la répression d'ERVs par H3K9me3 a également été observée dans des cellules germinales primaires et dans des cellules neurales progénitrices (227, 228).

e. Autres fonctions régulatrices

H3K9me3 ne serait pas uniquement impliquée dans la formation de l'hétérochromatine ou la répression génique. H3K9me3 peut également corrélér de façon positive avec la transcription. En effet, Vakoc et collaborateurs ont observé d'une part un enrichissement d'H3K9me3 sur le corps de gènes transcriptionnellement actifs (229). Ces auteurs ont notamment montré qu'après engagement du TCR, H3K9me3 s'accumule de façon concomitante à H3K4me3 sur le gène *Il2* dans les cellules T humaines. S'il s'agit d'un processus actif, puisque la marque n'est pas détectée sur ce locus avant activation, le rôle d'H3K9me3 en aval du TSS et sur le corps du gène n'est pas clairement établi. Des travaux plus récents, notamment réalisés par le groupe du Dr Muchardt à l'Institut Pasteur, suggèrent qu'H3K9me3 pourrait être impliquée dans la modulation qualitative de la transcription. Son accumulation sur des exons variants du gène *Cd44* corrèle avec le recrutement d'HP1 γ et avec un ralentissement de la vitesse d'élongation de l'ARN polymérase II, permettant au spliceosome d'inclure ces exons alternatifs au transcrit naissant (230). Ce rôle d'H3K9me3 dans l'épissage alternatif n'est pas sans rappeler celui de la méthylation de l'ADN également étudié sur le gène *Cd44* (188). Les deux marques pourraient coopérer dans ce processus. Les fonctions biologiques d'H3K9me3 sont donc multiples et dépendantes de la région génomique ciblée.

3. H3K9 Lysine méthyl transférase

Les KMT responsables de la méthylation d'H3K9 sont distinctes de celles responsables de la méthylation d'autres résidus. Elles appartiennent en grande majorité à la famille SUV39. Il s'agit de SUV39H1, SUV39H2, SETDB1, SETDB2, G9A et GLP, (231). Si SETDB1/2 et SUV39H1/2 ont été décrites comme pouvant tri-méthyliser H3K9 ce n'est pas le cas de G9A et GLP qui peuvent seulement di-méthyliser H3K9. Je centrerai mon propos sur les SETDB1/2 et SUV39H1/2. (Figure 20)

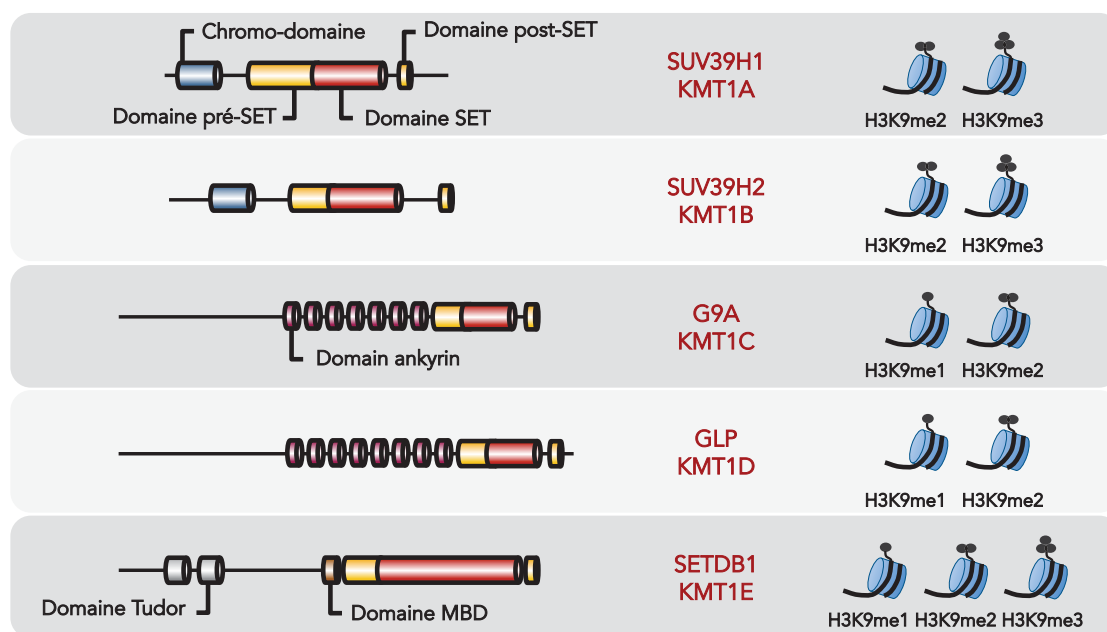


Figure 20 : Principales lysine méthyl transférase impliquées dans la méthylation d'H3K9

Le tableau présente les KMT de la famille SUV39h. Les principaux domaines fonctionnels des enzymes, le nom et la nomenclature de chaque KMT, et le nombre de groupements méthyle que chaque KMT peut transférer sur H3K9, sont indiqués. SET Suppressor of variegation Enhancer of zeste, Trithorax ; KMT Lysine (K) Methyl Transferase ;

Le rôle respectif des différentes enzymes susceptibles de tri-méthyliser H3K9 est difficile à appréhender. En effet, elles peuvent agir de façon indépendante ou coordonnée selon le contexte. La distribution de SUV39H1/2 et de SETDB1 a longtemps été considérée comme dichotomique. SUV39H1/2 étaient essentiellement décrites dans l'hétérochromatine (206, 210) et SETDB1 dans l'euchromatine (211). Cette distribution différentielle suggérerait une activité indépendante. Cependant SUV39H1 a maintenant été décrite dans l'euchromatine où elle

permet par exemple la répression du gène *Ifng* dans des LT de type Th2 via la déposition d'H3K9me3 sur le promoteur du gène (212). De plus, le groupe du Dr. Almouzni a suggéré l'existence d'un processus de méthylation d'H3K9me3 reposant sur l'action successive de SETDB1 et de SUV39H1 (232). Lorsque SETDB1 est associée au complexe formé de CAF-1 (Chromatin Assembly Factor 1) et d'HP1 α , SETDB1 mono-méthyle H3K9. H3K9me1 servirait alors de substrat à SUV39H1/2 pour générer H3K9me3 au niveau des régions péricentromériques lors de la réplication de l'ADN. Ce modèle est en accord avec la différence de substrat préférentielle de SETDB1 et de SUV39H1. En effet l'efficacité de l'activité méthyltransférase de SUV39H1 est inférieure de près de 50% en présence d'un substrat H3K9 non méthylé en comparaison à un substrat monométhylé. A l'inverse, la capacité de SETDB1 à triméthyliser H3K9 ne semble pas impactée par le niveau de méthylation du résidu (233). De plus, le groupe du Pr. Jenuwein a observé qu'en absence de SUV39H1, H3K9me3 disparaît des régions péricentromériques alors que H3K9me1 s'y accumule, confortant ainsi le modèle cité précédemment (206). Enfin, dans les mESC, le groupe du Dr. Jenuwein propose une autre forme de coopération entre SETDB1 et SUV39H1 (224). Afin de réprimer les ERVs, SETDB1 déposerait H3K9me3 sur le génome viral alors que SUV39H1 propagerait la marque sur les régions qui le flanquent.

Ayant principalement travaillé sur des souris déficientes pour SETDB1 au cours de ma thèse, je me focaliserai sur cette dernière dans la suite de ce manuscrit. Comme toutes les KMT de la famille SUV39, SETDB1 possède un domaine catalytique SET. Celui-ci a la caractéristique d'être « bifurqué », c'est-à-dire interrompu par une centaine d'acides aminés. L'impact fonctionnel de cette interruption n'est pas connu à ce jour. SETDB1 comprend également un domaine MBD lui permettant de se lier à l'ADN méthylé. Enfin, SETDB1 porte deux domaines tudors consécutifs ou TTD (Tandem Tudor Domain). Si la fonction du TTD n'est également pas connue, les domaines Tudors reconnaissent les arginines et lysines méthylées, notamment H3K9me3, et sont donc impliqués dans le ciblage de SETDB1 sur la chromatine. De plus, la présence des TTD ainsi que du MBD pourrait permettre de coordonner la méthylation de l'ADN à celle d'H3K9. En effet ces deux mécanismes de régulation épigénétique, bien que pouvant agir indépendamment et de façon non redondante (218), sont impliqués dans des fonctions biologiques similaires, que cela concerne la répression génique, l'épissage alternatif ou la répression des TE. Il a même été montré que les deux mécanismes agissent en synergie pour réprimer les IAP dans les mESC (219). En effet, dans des mESC déficientes en *Dnmt1*^{-/-} ou en *Setdb1*^{-/-}, certains IAP ne sont plus réprimés et sont exprimés. Ce niveau d'expression est plus que triplé dans les mESC double déficientes pour *Dnmt1*^{-/-} et *Setdb1*^{-/-}. Quoi qu'il en soit SETDB1, via la déposition d'H3K9me3, a un rôle clef et non redondant dans la répression de gènes et d'éléments transposables. Enfin si SETDB1

comprend une séquence d'import nucléaire lui permettant d'être localisée dans le noyau cellulaire, elle comprend également deux séquences d'export nucléaire (NES) lui permettant d'être localisée dans le cytoplasme. Il a par exemple été montré par le groupe du Dr Ait-Si-Ali que SETDB1 est relocalisée dans le cytoplasme lors de la différenciation terminale de myoblast pour y subir des MPT dont les fonctions restent à déterminer (234). Néanmoins la localisation cytoplasmique de SETDB1 et son impact biologique restent peu étudiés.

La répression transcriptionnelle de gènes et d'éléments transposables est une des fonctions majeures de SETDB1. Le principal modèle de son recrutement fait intervenir la famille des facteurs de transcription KZFP (KRAB domain and Zinc Finger Protein) et le corépresseur transcriptionnel TRIM28 (aussi appelé KAP-1) (211) (**Figure 21**). La succession de plusieurs domaines à doigts de zinc, dont le nombre est variable, permet au KZFP de se lier à une séquence d'ADN de façon hautement spécifique. Le corépresseur transcriptionnel TRIM28 est ensuite recruté via une interaction impliquant son domaine RBCC avec le domaine KRAB de la KZFP. Le domaine répresseur bi-partite de TRIM28 composé d'un domaine PH et d'un domaine Bromo va ensuite permettre le recrutement de SETDB1 via une interaction directe. La région de SETDB1 impliquée dans cette liaison est qualifiée de domaine KID (KAP-1 Interaction Domain). Plus récemment, il a été décrit que SETDB1 pouvait être ciblée dans la lignée de cellules cancéreuse mammaires T47D sur les promoteurs des gènes *AR* et *PR*, codant respectivement pour les récepteurs aux androgènes et à la progestérone. Dans ce modèle, la déposition de SETDB1 dépend d'événements moléculaires impliquant initialement le recrutement sur ces loci de la protéine Argonaute AGO2 par des ARN interférants (235).

4. H3K9me3 readers: L'interactome HP1

H3K9me3 est reconnue par les protéines HP1 (Heterochromatin Protein 1). Cette famille protéique est phylogénétiquement conservée et est composée de trois isoformes chez les mammifères (α , β et γ). Toutes les isoformes sont composées d'un domaine chromo (CD), d'une région charnière et d'un domaine chromo-shadow (CSD). Le CD permet aux protéines HP1 d'être recrutées sur H3K9me3 (236, 237). La région charnière peut aussi permettre aux protéines HP1 d'être ciblées sur la chromatine, notamment l'hétérochromatine, mais nécessite une interaction avec de l'ARN (238). Enfin le CSD permet aux protéines HP1 de recruter divers partenaires (**Figure 21**).

Si HP1 a tout d'abord été décrite comme étant impliquée dans la formation de l'hétérochromatine constitutive, on sait maintenant qu'elle peut également être enrichie dans l'euchromatine. Les fonctions respectives des différentes isoformes restent sujettes à investigation. Néanmoins HP1 α a fortement été associée à la formation de l'hétérochromatine dans des fibroblastes embryonnaires murins (MEF) (232). Plus récemment, elle a aussi été associée à la répression génique de l'*Ifng* dans des LT de type Th2 (212). Si HP1 α et HP1 β sont essentiellement associées à une conformation compactée de la chromatine, et donc à de la répression génique, les fonctions d'HP1 γ semblent plus diverses. En effet, il a été montré qu'elle s'accumule sur le corps de gènes induits après stimulation et qu'à l'inverse elle disparaît lors de leur répression (229). De plus, cette isoforme s'accumule sur H3K9me3 au niveau des exons alternatifs du gène *Cd44* permettant ainsi leur inclusion lors de l'épissage du transcrit naissant (230).

5. H3K9me3 déméthylases

Les membres de la famille KDM4 (JMJD2) sont capables de déméthyliser H3K9me3 (239) mais KDM4a et KDM4c peuvent également déméthyliser H3K36me2/me3. Leur expression est finement régulée et répond à des cinétiques précises lors des processus de différenciation. Les mécanismes moléculaires sous-jacents n'ont cependant pas encore été caractérisés. Les rôles biologiques de chacune de ces KDM restent sujets à investigation mais elles semblent non redondantes et leur activité pourrait être spécifique du type cellulaire. KDM4a, KDM4c et KDM4d ont toutes trois été impliquées dans la régulation de l'expression de gènes inductibles dans des contextes différents. Lors de la différenciation des mESC en cellules endothéliales sous l'influence du facteur de croissance BMP4, les deux isoformes KDM4a et KDM4c sont induites selon une cinétique différente et sont respectivement nécessaires à l'expression des gènes *Flk1* et *VE-cadherin* (240). Les deux enzymes sont essentielles à une différenciation appropriée des cellules endothéliales et à la vasculogénèse *in vivo*, tout en ciblant des gènes différents. Cependant cette étude n'a été menée que sur un set restreint de gènes. En ce qui concerne KDM4d (JMJD2d), son expression est induite dans les DC stimulées au LPS (216). Dans ce contexte, KDM4d permet de déméthyliser H3K9me3 sur les régions qui flanquent les enhancers des gènes inductibles *Mdc* et *Il12b* afin que ceux-ci soient exprimés. KDM4d est un régulateur inductible fort de l'expression génique. En effet, l'expression de *Mdc* et d'*Il12b* est fortement diminuée si les DC sont déficientes pour KDM4d. De plus, l'expression ectopique de KDM4d dans des fibroblastes 3T3 incapables d'exprimer les gènes inductibles *Mdc* et *Il12b*

induit la déméthylation d'H3K9me3 et permet l'expression ectopique de ces gènes après stimulation.

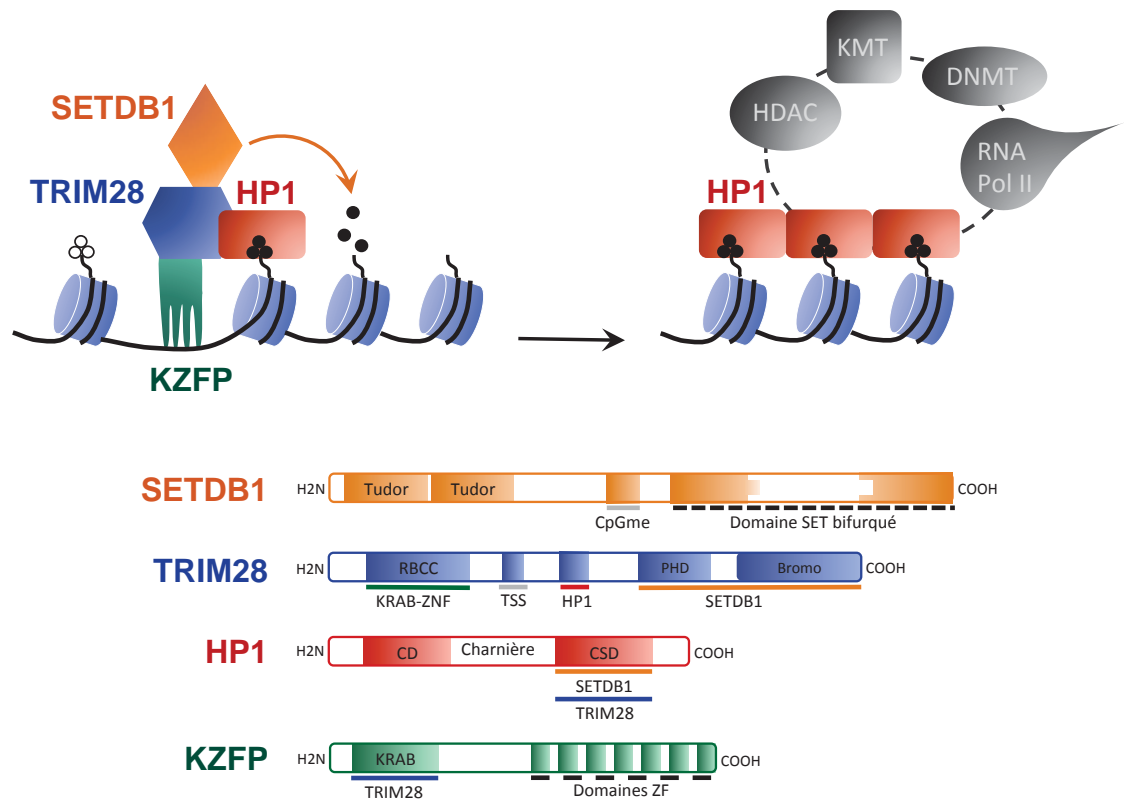


Figure 21 : Exemple d'interactome de SETDB1

Des protéines à doigts de zinc et domaines KRAB (KZFP) sont capables de reconnaître des séquences spécifiques de l'ADN grâce à une succession de plusieurs domaines en doigt de zinc (ZF). Leur domaine KRAB leur permet d'interagir avec TRIM28 qui peut interagir SETDB1 via ses domaines PHD et Bromo. HP1 peut interagir avec SETDB1 et TRIM28 via son domaine chromoshadow (CSD). La fixation d'HP1 sur H3K9me3 via le domaine chromo permet la propagation d'hétérochromatine locale.

Chapitre III : Régulation épigénétique de la biologie des lymphocytes T CD4

Le déterminisme cellulaire englobe deux mécanismes parallèles. D'une part : l'engagement d'une cellule dans un lignage, ou un stade de différenciation donné, qui se traduit par l'acquisition d'un phénotype et de fonctions caractéristiques. D'autre part : le verrouillage des phénotypes et des fonctions de lignages alternatifs. En effet une cellule engagée dans un lignage n'ayant pas effectué ces verrouillages restera très sensible aux signaux environnementaux. Le déterminisme cellulaire nécessite l'exclusion de potentiels pour une cellule. Ces acquisitions et pertes de fonctions se traduisent respectivement par une ouverture et une compaction de la chromatine. Ce remodelage se produit lors de la différenciation des thymocytes mais également lors de la différenciation des LT_N en LT helper. Il conditionne aussi la capacité de réponse des lymphocytes aux signaux activateurs ou répresseurs présents dans leur environnement lors d'une réponse immunitaire. Les régulations épigénétiques sont ainsi importantes pour le développement des LT dans le thymus et pour leur bon fonctionnement en périphérie.

I. Régulation épigénétique du développement intrathymique des lymphocytes T CD4

Les principaux acteurs moléculaires impliqués dans la régulation épigénétique du développement intrathymique des lymphocytes T CD4 et leurs fonctions sont résumés dans le **Tableau 1**.

SATB1 est une protéine qualifiée d'organisateur chromatinien. Elle permet d'établir des boucles de chromatine afin de réguler l'activité transcriptionnelle de régions génomiques. SATB1 est exprimée précocement lors de la différenciation des thymocytes mais est absente de la lignée myéloïde suggérant ainsi un rôle dans l'engagement des cellules dans le lignage lymphoïde. Les souris *Satb1*^{-/-} présentent un défaut de développement des LT dans le thymus dû à une réduction des capacités de lymphopoïèse par les cellules souches hématopoïétiques murines (mHSC). De plus l'expression ectopique de STAB1 dans des mESC induit la lymphopoïèse et la restaure dans des mHSC âgées qui n'expriment plus la protéine. SATB1 est

donc un des premiers acteurs de l'engagement des cellules progénitrices dans le lignage lymphoïde (241).

Durant leur développement dans le thymus, les thymocytes effectuent des choix séquentiels de lignage et la méthylation de l'ADN joue un rôle primordial dans la restriction de potentiels lignages alternatifs. Le groupe du Dr. C. Wilson a généré des modèles murins dans lequel *Dnmt1* est flanqué de séquence loxP et qui expriment la recombinase CRE sous le contrôle du promoteur, soit du gène *Lck*, soit du gène *Cd4* (242). Ces modèles permettent ainsi la délétion de *Dnmt1* au stade DN2 et DP respectivement. La déplétion de DNMT1 dans les thymocytes DP ne provoque pas de défaut majeur dans le thymus. En revanche, lorsque *Dnmt1* est supprimé précocement, les auteurs ont observé une diminution drastique de thymocytes DP accompagnée d'une nette augmentation du nombre de LT $\gamma\delta$. Ces derniers peuvent se développer dans le thymus à partir des thymocytes DN2 et DN3, et en l'absence de DNMT1, leur différenciation est biaisée vers le lignage T $\gamma\delta$. DNMT1 est donc importante pour restreindre les potentiels de différenciation des thymocytes et pour permettre le déterminisme dans le lignage T $\alpha\beta$.

Les HDAC1 et 2 sont nécessaires au développement des thymocytes SP CD4. Si la délétion de l'une ou de l'autre n'a pas d'effet sur le développement des thymocytes de par leurs fonctions partiellement redondantes, la délétion des deux gènes au stade de thymocytes DN provoque une accumulation de thymocytes immatures simples positifs qui précède le stade DP, révélant un arrêt du développement. De plus la réduction de la proportion de thymocytes DP TCR β^{high} CD69 $^{\text{high}}$ témoigne également d'un défaut de sélection positive. De façon cohérente, les auteurs observent une hausse d'H3K9ac en absence des HDAC1/2 qui se traduit par une dérégulation du transcriptome des thymocytes (243).

La survie des thymocytes DP et SP dépend de la lysine méthyl transférase SETDB1 spécifique d'H3K9 (244, 245). Il a été montré que la délétion de *Setdb1* dans les thymocytes DN induit l'apoptose des thymocytes DP et SP. L'absence de SETDB1 est associée à une surexpression du Fc γ RIIB de par l'absence d'H3K9me3 sur son promoteur. Ceci interfère par un mécanisme moléculaire non identifié avec la voie ERK en aval du TCR et provoque un défaut de survie. Néanmoins le défaut de signalisation de ERK en absence de SETDB1 est conséquent et ne peut être entièrement restauré par la délétion de *Fcgr2b*, suggérant l'implication de mécanismes complémentaires. Les rôles d'autres KMT spécifiques d'H3K9 ont été investigués mais les déplétions de G9a et SUV39H1 n'ont pas impacté le développement intra-thymique des LT (212, 246),

Le réarrangement de gènes codant pour le TCR est aussi sous l'influence de régulation épigénétique. Tout d'abord l'expression des gènes *Rag1* et *Rag2* est finement contrôlée. En amont de *Rag2* se trouve un élément « anti-silencer » (ASE) qui contrecarre l'activité d'un élément répresseur de leur expression (247). L'expression de l'organisateur chromatinien SATB1 atteint son niveau maximum dans les thymocytes DP. SATB1 se lie à l'ASE afin de former une boucle chromatinienne avec les promoteurs de *Rag1* et *Rag2* et permet le recrutement de l'ARN polymérase II. En absence de SATB1, *Rag1* et *Rag2* ne sont pas exprimés et le réarrangement des gènes codant pour le TCR α est altéré. Enfin après signalisation du TCR, les gènes *Rag1*, *Rag2* mais également *Dnmt* qui code pour la TdT sont réprimés. Il a été montré que lors de sa répression, ce dernier était relocalisé dans l'hétérochromatine à proximité des γ -satellites (248). De plus, les régulations épigénétiques affectent directement le locus du TCR α . En effet il a été montré que TRIM28 en se liant à l'enhancer E α des gènes codant pour le TCR α , permettait l'accumulation d'H3K4me2 sur des segments J α et le recrutement de RAG1 et/ou RAG2 (249). En l'absence de TRIM28 supprimé au stade DN, le réarrangement du TCR α est compromis et s'accompagne d'une accumulation de thymocytes DP TCR β^{low} et d'un défaut de thymocytes SP CD4.

Le groupe de Dr. Trono a étudié le rôle de TRIM28 à l'échelle du génome dans le développement des lymphocytes T avec un modèle murin similaire (250). Si le défaut de réarrangement du TCR α n'a été ni confirmé ni infirmé, la proportion de thymocytes SP CD4, mais pas celle des thymocytes SP CD8, est bien réduite en absence de TRIM28. Des analyses transcriptomiques ont révélé que la délétion de TRIM28 dans les thymocytes DN déréprime des gènes impliqués dans l'activation du TCR, la signalisation des cytokines, la survie cellulaire et le réarrangement du cytosquelette. La plupart d'entre eux sont directement ciblés par TRIM28, ce qui permet la déposition d'H3K9me3 en condition sauvage et leur répression. Ainsi TRIM28 est impliqué dans de multiples étapes du développement des lymphocytes T allant de la régulation du réarrangement du TCR α à la signalisation en passant par le choix de lignage CD4 ou CD8.

Le domaine PxVxL de TRIM28 lui permet d'interagir avec les protéines HP1 impliquées dans la reconnaissance d'H3K9me3. Il est donc cohérent que la stabilité du choix de lignage CD4 ou CD8 soit également dépendante de HP1 β , comme suggéré par le groupe de Dr. Littman. En effet il a mis en évidence un élément intra-génique répresseur du gène *Cd4* (S4) nécessaire à la répression du corécepteur dans les thymocytes DP et SP CD8 chez la souris (251). Ce répresseur est composé de plusieurs sites aux fonctions partiellement redondantes de sorte que leur délétion ponctuelle n'entraîne qu'une perte partielle de la répression. La surexpression d'HP1 β dans ce contexte restaure partiellement la répression, suggérant ainsi

que des KMT non identifiées à ce jour pourraient enrichir l'élément répresseur en H3K9me3 afin de verrouiller *Cd4*. Il est envisageable que dans ce contexte, TRIM28 recrute SETDB1 qui va déposer H3K9me3 permettant ainsi la compaction du locus par HP1 β mais ce modèle n'a pas été testé expérimentalement. De plus il a été montré que le locus *Cd4* est enrichi en 5mC dans les thymocytes DN3; la préservation de ce patron de méthylation par DNMT1 dans les thymocytes SP CD8 nécessite la présence de S4 (252). A l'inverse, la présence de l'enhancer (E4_P) est nécessaire à la déméthylation du locus dans les thymocytes CD4 SP. Le choix de lignage CD4 ou CD8 implique donc de nombreux acteurs épigénétiques et met en jeux des voies répressives complémentaires. Au cours de sa répression, le gène codant pour le co-récepteur réprimé est relocalisé dans l'hétérochromatine (253).

Un autre choix de lignage s'effectue dans le thymus lors du développement des lymphocytes T CD4. Il s'agit du choix entre le lignage T régulateur et T conventionnel. SATB1 fonctionne comme un facteur pionnier de la différenciation des LT régulateurs thymiques car il est responsable de l'activation de super-enhancers du lignage T régulateur thymique (254). Dans les thymocytes DP SATB1 s'accumule sur les régions non codantes conservées de *Foxp3*. Au cours de la différenciation en thymocytes régulateurs, ces mêmes régions accumulent des marques permissives comme H3K4me1 et H3K27ac. En absence de SATB1, ces *loci* restent réprimés et les précurseurs thymiques perdent leur potentiel d'engagement dans le lignage T régulateur. A l'inverse, des acteurs épigénétiques sont impliqués dans la répression du programme T régulateur dans les thymocytes SP CD4 n'exprimant pas *Foxp3*. C'est le cas de la SUMO-E3 ligase PIAS1 (Protein Inhibitor of the Activated STAT1) qui maintient la chromatine au niveau du promoteur de *Foxp3* dans une conformation répressive (255). Le promoteur est en effet enrichi en 5mC, H3K9me3 et HP1 γ . Si la KMT responsable de la déposition d'H3K9me3 n'a pas été identifiée dans ce contexte, il a été montré que PIAS1 est nécessaire pour le recrutement de DNMT3A/B sur le promoteur de *Foxp3*. En absence de PIAS1, le locus est déméthylé et H3K9me3 n'est plus enrichi, ce qui se traduit par une augmentation de la proportion de thymocytes *Foxp3*⁺ et une résistance des souris *Pias1*^{fl/fl}x*Cd4*CRE à l'induction de l'EAE (encéphalite auto-immune expérimentale, un modèle murin de sclérose en plaque).

Protéine	Fonctions biologiques	Réf.
SATB1	<ul style="list-style-type: none"> • Favorise l'engagement des mHCS dans la lignée lymphoïde • Nécessaire à l'expression de <i>Rag1</i> et <i>Rag2</i> et donc au réarrangement des gènes codant le TCR • Nécessaire à l'engagement dans le lignage T régulateur 	(240) (247) (253)
DNMT1	• Favorise l'engagement des thymocytes dans le lignage T $\alpha\beta$	(241)
DNMT3	• Impliquée dans la répression du lignage T régulateur	(254)
HDAC1/2	• Nécessaires au développement des thymocytes SP CD4 et à leur sélection positive	(242)
SETDB1	• Nécessaire à la survie des thymocytes DP et SP	(243-244)
TRIM28	<ul style="list-style-type: none"> • Nécessaire au réarrangement des gènes codant pour le TCRα • Nécessaire à l'engagement des thymocytes dans le lignage CD4 	(248) (249)
PIAS1	• Impliquée dans la répression du programme T régulateur dans les thymocytes SP CD4	(254)
HP1 β	• Impliquée dans la stabilité des lignages CD4 et CD8	(250)
HP1 γ	• Impliquée dans la stabilité des lignages T régulateurs et conventionnels	(254)

Tableau 1: Principaux acteurs moléculaires impliqués dans la régulation épigénétique du développement intrathymique des lymphocytes T CD4 et leurs fonctions

II. Régulation épigénétique des lymphocytes T CD4 en périphérie

Dans les lymphocytes T naïfs, les régions promotrices de la plupart des gènes caractéristiques des différents lignages T helper, comme *Tbx21* et *Gata3* qui codent pour les « master regulators » des lignages Th1 et Th2, sont dans un état « poised ». Elles sont en effet enrichies à la fois en marques permissives (H3K4me3, H3ac) et répressives (H3K27me3, 5mC...) (256) ce qui leur permet d'être très rapidement mobilisées, ou au contraire verrouillées, en fonction des signaux environnementaux reçus par la cellule. La chromatine va être remodelée lorsque les LT_N reçoivent les signaux polarisants. Ce remodelage comprend la déposition de marques permissives sur les *loci* spécifiques du lignage induit, et de façon concomitante, la déposition de marques répressives sur les *loci* des devenir alternatifs (**Figure 22**). Par exemple, dans le cas de lymphocytes T qui se différencient en lymphocytes Th1, H3K27me3 disparaît du promoteur de *Tbx21* alors que H3K4me3 s'accumule. A l'inverse, lors de la différenciation de lymphocytes T vers un phénotype Th2, le signal H3K4me3 n'augmente pas sur le promoteur de ce gène et le complexe PRC2 est recruté sur ce locus. Ce schéma n'est cependant pas systématique, ce qui explique notamment par exemple la différence de plasticité des cellules Th2 et Th17 vers le lignage Th1. Les régulations épigénétiques conditionnent donc la réponse des LT CD4 aux signaux environnementaux, qu'ils s'agissent des signaux de différenciation ou de survie. Lorsque l'épigénome est altéré, la réponse du LT sera ainsi aberrante. Illustrant le rôle causal des marques épigénétiques dans la programmation des LT, des cellules naïves traitées avec du 5-Aza-2-deoxycytidine, un analogue de base qui s'incorpore à l'ADN à la place des cytosines et qui réprime également les DNMT, inhibant ainsi la méthylation de l'ADN ; ou du sodium butyrate, un inhibiteur des HDAC, produisent de façon aberrante de l'IFN- γ et de l'IL-4 (68).

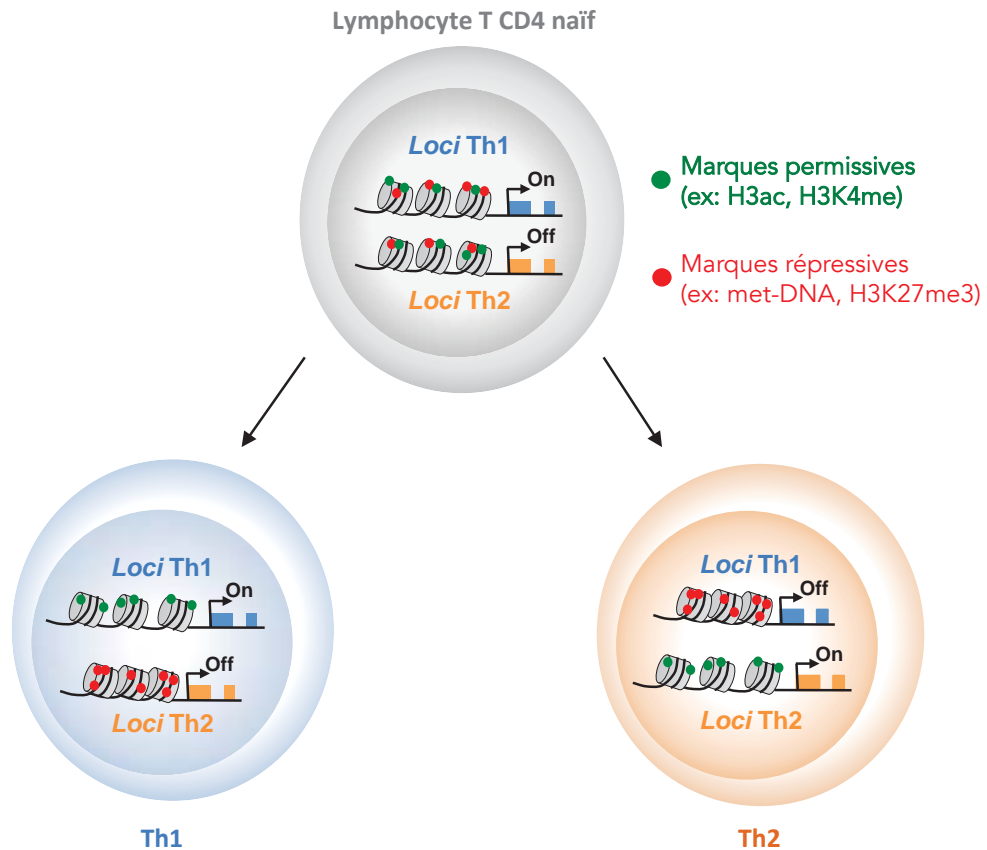


Figure 22 : La différenciation des lymphocytes T CD4 naïfs s'accompagne d'un remodelage de la chromatine

Les *loci* caractéristiques des programmes d'expression génique Th1 et Th2 sont « poised » dans les lymphocytes T CD4 naïfs. Ils portent à la fois des marques permissives et répressives. Lors de la différenciation, un remodelage de la chromatine s'opère sous le contrôle des facteurs de transcription pionniers. Ce remodelage comprend la déposition de marques permissives sur les *loci* spécifiques du lignage induit, et répressives sur les *loci* des devenir alternatifs.

A. Régulation épigénétique des fonctions effectrices

1. Remodelage de la chromatine par les facteurs pionniers

a. Activation des enhancers par les STAT

Ces dix dernières années, les techniques d'analyse à l'échelle du génome ont permis de cartographier la localisation et l'état d'activation des enhancers propres à chaque lignage de lymphocytes T helper. Le répertoire d'enhancers de chaque lignage est unique. Pendant longtemps, il a été considéré que les facteurs de transcription master régulateurs coordonnaient l'engagement des cellules dans les lignages, et donc la mobilisation des réseaux d'enhancers qui contrôlent les programmes d'expression génique qui leurs sont propres. Diverses études ont cependant montré que cette fonction incombait essentiellement aux STATs et à d'autres facteurs de transcription pionniers (62, 257, 258). Ciofani et collaborateurs ont par exemple montré que les facteurs de transcription BATF et IRF4 (activés en aval de la signalisation TCR) et STAT3 (activé en réponse à la signalisation de l'IL-6) étaient responsables de l'induction du patron d'enhancers spécifique du lignage Th17 (257). La délétion d'un de ces facteurs induit une diminution importante de la liaison de p300 servant d'indicateur de l'activité des enhancers alors que celle de RoRyT n'a que peu d'effet. RoRyT semble en revanche important pour moduler finement (positivement ou négativement) l'expression des gènes contrôlés par ces enhancers. De plus Vahedi et collaborateurs ont montré dans un article référence que STAT4 et STAT1 d'une part, et STAT6 d'autre part, étaient respectivement responsables de l'activation de 60% et 77% des enhancers Th1 et Th2 (62). Confirmant le rôle mineur des « master regulators » dans ce processus, la surexpression de TBET et GATA3 ne permet de restaurer que très partiellement le patron d'enhancers actifs Th1 et Th2. L'importance de STAT6 dans l'activation des enhancers de type Th2 a également été soulignée chez l'Homme (258). De plus, il a été montré que STAT6 est rapidement recrutée en condition polarisante Th2 aux niveaux d'enhancers caractéristiques du lignage. En revanche, l'acétylation d'H3K27, et donc l'activation des enhancers, requiert plus de temps ce qui est cohérent avec le fait que STAT6 est un facteur initiateur de la différenciation Th2.

Les STATs et autres facteurs pionniers sont donc responsables du façonnage des patrons d'enhancers actifs caractéristiques des lignages de T helper. En effet, les facteurs pionniers

permettent de cibler des « chromatin writers » qui vont déposer des marques permissives sur les enhancers du lignage. Le façonnage de ce patron repose également sur d'autres mécanismes comme par exemple la réorganisation spatiale de la chromatine ou encore l'exclusion de complexes répresseurs que j'aborde dans les sections suivantes (**Figure 25**).

b. Déplacement des nucléosomes et réorganisation spatiale de la chromatine

Il a été montré que STAT4 et NFAT permettent le recrutement précoce de BRG1 (Brahma Related Gene 1) sur le locus *Ifng* au cours de la différenciation Th1 (259). BRG1 permet de déplacer les nucléosomes au niveau des enhancers et du promoteur du gène (**Figure 23**). Ce remodelage est nécessaire à l'expression génique. En effet, la délétion de BRG1 provoque une nette diminution de la proportion de cellule productrice d'IFN γ . De façon similaire, BRG1 est recrutée par STAT6 et NFAT dans les cellules Th2 au niveau de l'enhancer RHS7 (Regulatory High Sensitivity 7) du locus Th2 (260).

Il a également été suggéré que les STATs soient à la source d'une réorganisation spatiale plus globale de la chromatine. Certaines régions de la chromatine vont avoir tendance à interagir ensemble; à l'inverse, d'autres régions peuvent être exclues d'un domaine chromatinien. L'interactome chromatinien du *Ifng* est distinct entre un LT_N, une cellule Th1 et un LT de type Th2 (261). Dans des lymphocytes Th1 déficients pour *Stat4*, l'interactome chromatinien de *Ifng* est altéré et présente de nombreuses similitudes avec celui des LT_N. Cette altération n'a pas d'impact sur la transcription des gènes associés mais l'état de la chromatine n'a pas été analysé. Il serait intéressant de savoir si cette relocalisation s'accompagne d'une modification des marques épigénétiques rendant les cellules plus sensibles à des signaux instructeurs d'autres lignages par exemple.

Enfin il a été montré que SATB1, à travers une réorganisation de la chromatine, était nécessaire à l'expression des cytokines Th2 par une lignée lymphocytaire de type Th2 (262). Les cellules de cette lignée expriment de façon constitutive GATA3 et produisent les cytokines Th2 lorsqu'elles sont stimulées par la concanavaline A. Ce modèle artificiel, a permis de mettre en évidence que SATB1 est nécessaire à l'organisation de boucles de chromatine actives enrichies en H3K9ac, ARN polymérase II, STAT6 et BRG1 (**Figure 23**). Cependant, si la délétion de SATB1 entraîne une diminution de la transcription des gènes codant pour les cytokines de type Th2, les auteurs n'ont pas déterminé l'ordre du déroulement de ces changements épigénétiques au niveau du locus codant les cytokines. De plus le groupe du Dr.

Flavell a mis en évidence que GATA3 et STAT6 sont essentiels pour l'établissement et la maintenance de cette boucle chromatinienne (153).

Toutes ces études permettent donc d'imaginer le modèle suivant : les facteurs de transcription pionniers recruteraient BRG1 pour déplacer les nucléosomes et SATB1 pour établir des boucles de chromatine sur les *loci* codant les cytokines du lignage. Le déplacement des nucléosomes et les boucles de chromatines faciliteraient alors la déposition de marques permissive et l'activation des *loci*.

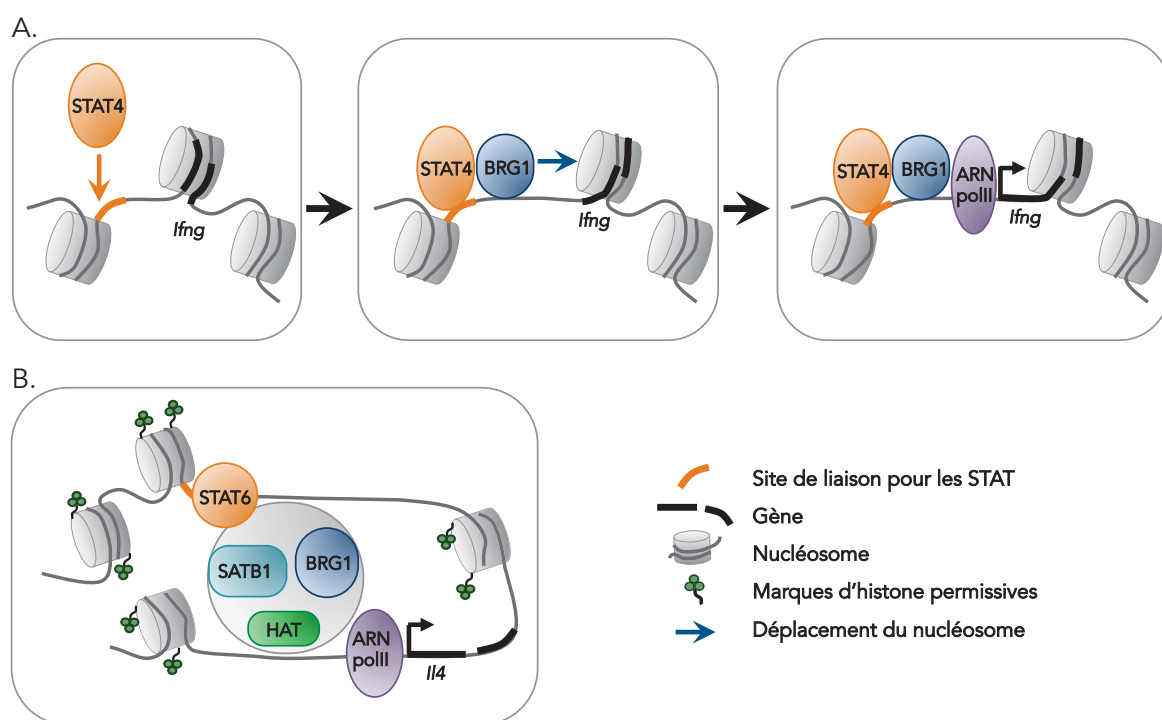


Figure 23 : Modèle de déplacement des nucléosomes et réorganisation de la chromatine

Les facteurs de transcription pionniers sont impliqués dans le recrutement de BRG1 et SATB1 qui permettent le déplacement des nucléosomes et la formation de boucle chromatinienne respectivement. Cette réorganisation de la chromatine faciliterait la déposition de marques épigénétiques permissives de la transcription. A. STAT4 permet le recrutement précoce de BRG1 sur le locus *Ifng* au cours de la différenciation Th1. BRG1 est impliqué dans le déplacement des nucléosome nécessaire à l'expression du gène. B. Dans une lignée lymphocytaire de type Th2, SATB1 est nécessaire à l'organisation de boucles de chromatine actives enrichies en marques d'histone permissives, ARN polymérase II, STAT6 et BRG1. SATB1 favorise ainsi la transcription des gènes codant pour les cytokines de type Th2.

c. Exclusion du complexe Polycomb

Les voies Polycomb (PcG) et Trithorax (TrxG) sont antagonistes. La première permet la déposition d'H3K27me3 et est donc répressive, tandis que la seconde permet la méthylation d'H3K4 et est donc permissive. Leur antagonisme est essentiel à la programmation des LT. En effet, lors de la différenciation des lymphocytes Th2, STAT-6 recrute la protéine Menin sur *Gata3*. Menin fait partie du complexe multimoléculaire TrxG dont elle permet la liaison à l'ADN (263). TrxG méthyle H3K4 mais permet également l'exclusion du complexe PcG. Le locus *Gata3* n'est ainsi plus réprimé et l'expression du facteur de transcription peut être soutenue. De plus, dans les cellules Th9, les facteurs de transcription Smad2 et Smad4 en aval de la signalisation du TGF- β permettent également de déplacer EZH2, la méthyltransférase du complexe PRC2 (264). Les facteurs de transcription pionniers déterminent donc le set d'enhancer actifs d'un lignage de lymphocytes T notamment en excluant des complexes répresseur de la chromatine.

2. La dé-méthylation de l'ADN

La découverte des ADN déméthylases de la famille TET a amené la communauté scientifique à s'interroger sur leur rôle dans la biologie des LT. La distribution à l'échelle du génome du 5hmC, un intermédiaire de dégradation des 5mC, a récemment été analysée dans des lymphocytes T naïfs et différentes populations de lymphocytes T helper. La distribution des 5hmC est spécifique du type cellulaire. Elles sont enrichies au niveau d'éléments régulateurs de gènes caractéristiques d'un lignage de lymphocytes T helper (265). En effet, dans des lymphocytes T de type Th1 et Th17, TET2 est ciblée sur les éléments régulateurs de *Ilfn3* et de *Il17* par Tbet et RoR γ T (**Figure 24**). La dégradation des 5mC par TET2 favorise le recrutement de l'histone acétyltransférase p300 et l'accumulation de marques épigénétiques permissives sur la chromatine. Dans les lymphocytes T régulateurs, TET1 et TET2 sont ciblées sur les éléments régulateurs de *Foxp3* par SMAD3 et STAT5 en aval de la signalisation de TGF- β et de l'IL-2 respectivement. Elles oxydent les 5mC afin de générer des 5hmC. En absence de TET1/2, *Foxp3* est réprimé par la présence de 5mC et les lymphocytes T régulateurs ne sont pas fonctionnels.

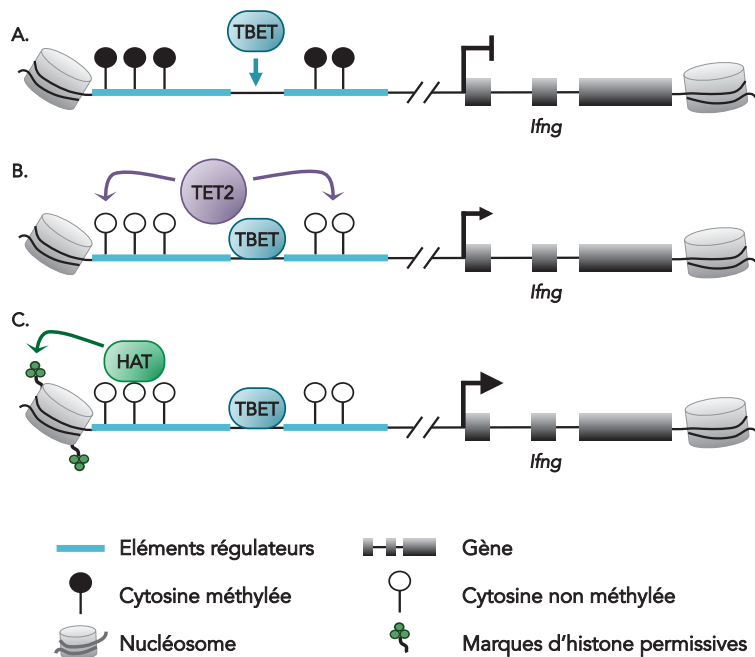


Figure 24 : La dé-méthylation des cytosines par les protéines TET est nécessaire à l'acquisition des fonctions effectrices par les lymphocytes T helper

A. Les régions régulatrices de l'*Ifng* sont méthylées dans les lymphocytes T naïfs, le gène *Ifng* n'est pas exprimé. Lors de la différenciation en lymphocytes T de type Th1, le facteur de transcription TBET se fixe au niveau de régions régulatrices. B. TBET permet le recrutement de l'ADN dé-méthylase TET2 qui hydrolyse le groupement méthyl des cytosines. C. Les cytosines non méthylées favorisent alors le recrutement d'histone acétylase (HAT) permettant l'accumulation de marques d'histones permissives. *In fine*, la dé-méthylation des cytosines sur les régions régulatrices de l'*Ifng* permet son expression génique.

Si la déméthylation de l'ADN est nécessaire à l'acquisition des fonctions effectrices par les LT helper, elle a également été impliquée dans la plasticité cellulaire. En effet, dans des conditions permettant la transdifférenciation de lymphocytes T Th2 vers le lignage Th1, il a été montré que STAT4 et Tbet induisent la diminution du niveau de méthylation de l'ADN au niveau des éléments régulateurs de l'*Ifng* (266). Néanmoins, dans cette étude aucune DNMT n'a été « incriminée ».

Des exemples de remodelage de la chromatine induit par les facteurs de transcription pionniers sont illustrés dans la **Figure 25**.

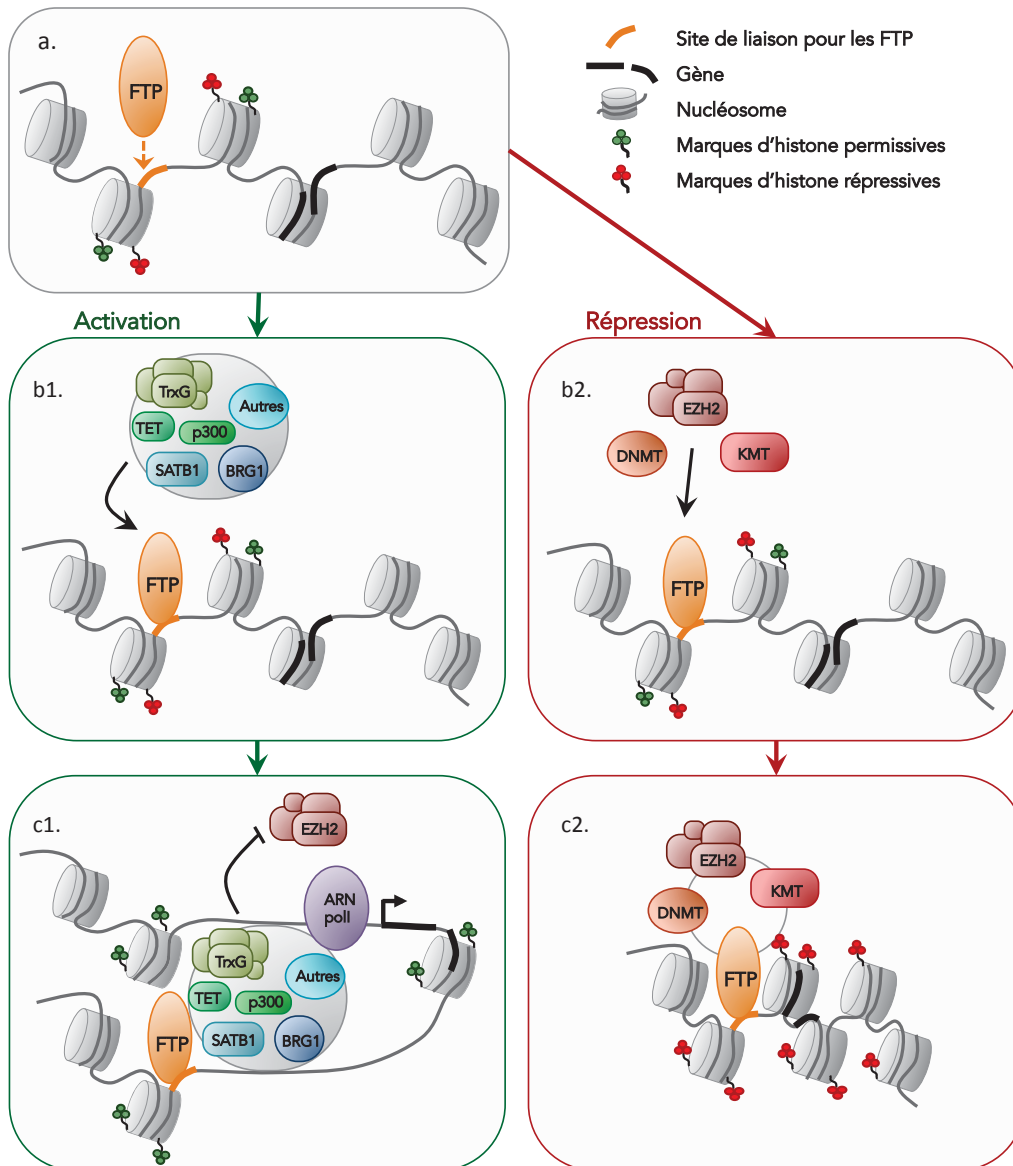


Figure 25 : Modèle de remodelage permissif et répressif de la chromatine

a. En réponse à divers stimuli, les facteurs de transcription pionniers (FTP) s'activent et se fixent sur leur motifs de liaison sur l'ADN. Ils permettent le recrutement direct ou non de divers facteurs impliqués dans le remodelage de la chromatine. La séquence des événements moléculaires n'étant pas encore bien définie, ces facteurs sont ici illustrés regroupés. b1. Sur les *loci* spécifiques du lignage induit, les FTP vont recruter diverses protéines impliquées dans l'ouverture de la chromatine. c1. Parmi ces protéines, SATB1 est par exemple impliquée dans la formation de boucle de chromatine et BRG1 dans le repositionnement des nucléosomes. Des chromatinines « witters » et « erasers » sont également recrutés afin de déposer des marques d'histone permissives et d'enlever des marques répressives respectivement. Leur présence peut également inhiber celle de complexes répresseurs. b2. Les FTP peuvent également être impliqués dans le recrutement de complexes répresseurs de la chromatine sur les *loci* des lignages alternatifs. c2. Ces complexes induisent l'accumulation de marques répressives et la compaction de la chromatine.

B. Répression épigénétique des lignages alternatifs

Comme expliqué dans le chapitre I, le paradigme Th1/Th2 reposait sur l'idée que les lymphocytes T de type Th1 et Th2 étaient des cellules stables en stade terminal de différenciation. Si ce paradigme est aujourd'hui remis en question pour les lignages Th17 ou T_{FH} par exemple, il existe néanmoins des mécanismes de verrouillage des programmes d'expression génique caractéristiques des lignages alternatifs au niveau de la chromatine nécessaire au bon déterminisme cellulaire (**Figure 25**).

1. Méthylation de l'ADN

DNMT1 a été impliquée dans la répression des cytokines de type Th2 et Th1 dans de nombreux lignages (242, 267, 268). Conformément à son rôle de DNMT de maintenance, DNMT1 semble essentielle pour maintenir le méthylome entre le stade de thymocytes SP CD4 et les LT_N. La DNMT3A *de novo* est quant à elle impliquée dans la répression de l'*Ifng* dans les lignages de lymphocytes T autres que Th1 (269). Si le promoteur du gène est dépourvu de 5mC dans des cellules T naïves et Th1, DNMT3A le méthyle dans les lymphocytes Th2, Th17 et iTreg. L'absence de DNMT3A n'a pas d'impact direct sur la différenciation et le phénotype des cellules Th2, Th17 et iTreg. En revanche, la perte de la méthylation au niveau du promoteur de l'*Ifng* altère leur stabilité. En effet, s'ils sont cultivés en condition polarisante de type Th1, ces LT différenciés se mettent à produire de l'IFN γ de façon aberrante.

2. La voie Polycomb – H3K27me3

EZH2 tri-méthyle H3K27 sur les promoteurs des gènes codant pour les cytokines et facteurs de transcription caractéristiques des lignages Th1 et Th2 dans les LT_N (270, 271). Au cours de la différenciation des LT_N, H3K27me3 est maintenue par EZH2 sur les gènes des lignages alternatifs (270, 272). La délétion d'EZH2 se traduit par une perte d'H3K27me3 et la surexpression de gènes caractéristiques de lignages alternatifs, mais surtout par une perte de détermination cellulaire. Par exemple, des lymphocytes Th1 déficients pour EZH2 cultivés en présence de signaux polarisant pro-Th2 produisent simultanément des cytokines de type Th1 et Th2 contrairement à des LT de type Th1 sauvage qui ne produisent que de l'IFN γ . Il est intéressant de noter qu'un mécanisme de ciblage d'EZH2 spécifique de lymphocytes Th1 a été mis en évidence chez l'Homme (273). L'ARN long non codant linc-MAF-4 est en effet exprimé

de façon spécifique dans les lymphocytes T Th1 dans lesquels il permet la répression du facteur de transcription pro-Th2 MAF-4. En l'absence du linc-MAF-4, les lymphocytes Th1 expriment des gènes caractéristiques des cellules Th2. Les auteurs proposent un modèle dans lequel EZH2 est recrutée au niveau du transcrit de *linc-MAF-4*; une boucle chromatinienne permet ensuite le rapprochement d'EZH2 au niveau du gène *Maf4* pour le réprimer dans les lymphocytes Th1.

Enfin, dans les lymphocytes T régulateurs, les groupes des Drs Rudensky et Bluestone ont montré que le « master regulator » FOXP3 recrute EZH2 pour trimétyler H3K27 sur des *loci* à réprimer (274, 275). Les auteurs suggèrent que ce mécanisme permet de réprimer les gènes, en aval de la signalisation du TCR et des molécules de co-stimulation, qui pourraient détourner les lymphocytes T de leurs fonctions régulatrices vers un rôle effecteur. La répression de gène via la tri-méthylation d'H3K27 par EZH2 est donc un mécanisme de verrouillage des lignages alternatifs et un garant de la stabilité cellulaire commun à différents lignages de lymphocyte T CD4.

3. Relocalisation des gènes

Le groupe du Dr. Locksley a montré par FISH que les gènes *Ifng* et *Il4* étaient distants des γ -satellites marquant l'hétérochromatine dans les lymphocytes T naïfs. En revanche les auteurs ont observé leur relocalisation au niveau de l'hétérochromatine dans les lymphocytes T de type Th2 et Th1 respectivement (69). Cette relocalisation est cohérente avec la perte de potentiel de différenciation observée à partir de cinq cycles de division lors des tests de différenciation in vitro. Cependant ces résultats sont controversés. En effet le groupe du Dr. Merckenschlager a montré que les gènes codant pour les cytokines ne sont pas relocalisés au niveau de l'hétérochromatine (276). En revanche *Gata3* et *c-maf* sont relocalisés au niveau de l'hétérochromatine centromérique présente à la périphérie nucléaire dans des lymphocytes Th1. Mais la relocalisation des gènes codant pour les facteurs de transcription des lignages alternatifs n'est pas un mécanisme ubiquitaire. En effet, dans les lymphocytes Th2, *Tbx21* n'est pas relocalisé au niveau de l'hétérochromatine.

C. Rôle de l'interactome d'H3K9me3 dans la biologie des LT

G9a est impliquée dans l'expression des cytokines dans les LT Th2 de façon indépendante de son activité méthyltransférase (246). En effet, bien qu'en son absence la production d'IL-13 soit diminuée, si les lymphocytes T naïfs ou de type Th2 sont traités avec un inhibiteur spécifique de son activité enzymatique, la production d'IL-13 n'est pas altérée. En revanche l'activité méthyltransférase de G9a est nécessaire pour réprimer l'expression de *Il17a* dans les LT_N et Th2.

La répression de l'IL-17 dans des LT de type Th17 en réponse à des signaux inhibiteurs, requiert la coopération de G9a et SETDB1 (277). En effet Xiao et collaborateurs ont montré que la production d'IL-17 par les LT Th17 était réprimée lorsque les LT interagissent avec des cellules dendritiques exprimant de façon constitutive le ligand de la molécule de co-stimulation OX40. La répression de *Il17a* par la signalisation d'OX40 est dépendante de RelB. RelB permet le recrutement de G9a et de SETDB1 sur les éléments κ BEBE (κ B Binding Element) du locus *Il17*. H3K9me2 et H3K9me3 s'accumulent alors et répriment par propagation les éléments de réponse à RoR γ T. La fixation du master régulateur est notamment réprimée et *Il17* ne peut être exprimé.

H3K9me3 a également été impliquée dans la répression du locus *Il2*. Dans les lymphocytes T régulateurs Foxp3+, H3K9me3 et le corépresseur TRIM28 s'accumulent sur les régions promotrices de ce gène (278) et leur délétion provoque une diminution de la méthylation d'H3K9 et une accumulation de l'acétylation du gène permettant son expression. Si la KMT responsable de la méthylation d'H3K9 n'a pas été identifiée dans ce contexte, on peut néanmoins émettre l'hypothèse qu'il s'agit de SETDB1 car TRIM28 en est un partenaire privilégié. En revanche, la répression de *Il2* par déposition d'H3K9me3 sur son promoteur proximal en réponse à la signalisation du TGF- β implique, quant à elle, les lysine méthyltransférases SUV39H1 et SUV39H2 (279).

Enfin, le groupe de Sébastien Amigorena a impliqué SUV39H1 et HP1 α dans la stabilité du lignage Th2 (212). Si des LT_N déficients pour SUV39H1 se différencient correctement en LT de type Th2, ils présentent en effet cependant un défaut de stabilité. En effet, lorsque les LT Th2 KO pour SUV39H1 sont cultivés en condition polarisante Th1, ils produisent de l'IFN γ contrairement à leur contrepartie sauvage. Cette plasticité s'explique par la perte partielle d'H3K9me3 et la diminution du recrutement d'HP1 α sur le promoteur de *Ilfng*. Néanmoins, la persistance d'H3K9me3 sur les enhancers de *Ilfng* et de *Tbx21* en absence de SUV39H1

suggère qu'une autre lysine méthyltransférase est impliquée dans la répression du programme Th1 dans les LT Th2.

Si l'interactome d'H3K9me3 a essentiellement été impliqué dans la répression génique dans les lymphocytes T, il semble néanmoins que son rôle soit plus complexe. De façon surprenante, il a été montré que STAT3 permet le recrutement de TRIM28 sur les super-enhancers dans les lymphocytes T de type Th17 afin de les activer. Il a été proposé que TRIM28 et l'histone acétyltransférase p300, en agissant de concert, permettent de clustériser ensemble les enhancers du gène codant pour l'IL-17. Ils sont de plus nécessaires au recrutement de RoR γ T sur ces régions (280). Néanmoins, le niveau transcriptionnel global est faiblement affecté par la délétion de TRIM28. Les fonctions activatrices de TRIM28, initialement décrit comme un co-répresseur, restent troubles et seront certainement mieux caractérisées dans le futur.

Comme je viens de le décrire, la fonction de la marque d'histone d'H3K9me3 dans la régulation de l'expression des gènes est complexe et reste encore globalement peu connue, notamment dans la biologie des lymphocytes T. C'est dans ce contexte que se situe ma thèse dont l'objectif était de caractériser le rôle de la lysine méthyl transférase SETDB1 une des enzymes majeures responsable du dépôt de cette marque.

Protéine	Cellules	Effet biologique	Réf.
G9a	LT _N	• Répression de l' <i>Il17a</i>	(245)
	TH2	• Répression de l' <i>Il17a</i>	(245)
		• Favorise l'expression de cytokines Th2 de façon indépendante de son activité méthyl transférase	(245)
	Th17	• Répression de l' <i>Il17a</i> par déposition d'H3K9me3 sur les éléments κ B _{BE} , en coopération avec SETDB1, en réponse à la signalisation d'OX40	(276)
SETDB1	Th17	• Répression de l' <i>Il17a</i> par déposition d'H3K9me3 sur les éléments κ B _{BE} , en coopération avec G9a, en réponse à la signalisation d'OX40	(276)
SUV39h1	Treg	• Répression de l' <i>Il2</i> par déposition d'H3K9me3 sur le promoteur	(278)
	Th2	• Répression de l' <i>Ifng</i> par déposition d'H3K9me3 sur le promoteur	(211)
SUV39h2	Treg	• Répression de l' <i>Il2</i> par déposition d'H3K9me3 sur le promoteur	(278)
TRIM28	Treg	• Répression de l' <i>Il2</i> associé à une accumulation d'H3K9me3 sur les régions promotrices	(277)
HP1α	Th2	• Répression de l' <i>Ifng</i> par déposition d'H3K9me3 sur le promoteur	(211)

Tableau 2: Rôle de l'interactome d'H3K9me3 dans la biologie des lymphocytes T CD4

CONTRIBUTIONS

Mon projet de thèse, débuté lors de mon stage de M2R, s'inscrit dans une thématique globale de l'équipe qui vise à étudier le rôle de l'interactome d'H3K9me3 dans la régulation épigénétique des lymphocytes T. La transversalité de ce projet a nécessité de mutualiser des compétences en épigénétique, bio-informatique et immunologie. Les travaux que je présente dans ce manuscrit de thèse sont le fruit d'un travail d'équipe dirigé par le Dr. Olivier Joffre et mené de front par le Dr. Véronique Adoue, le Dr. Bénédicte Binet (alors étudiante en thèse), Joanna Fourquet (IE bioinformatique) et moi-même.

Le projet principal du groupe a porté sur l'étude du rôle de la lysine méthyl transférase SETDB1 dans la biologie des lymphocytes T CD4. Lors de mon stage de M2R j'ai commencé à caractériser le phénotype des souris déficientes pour SETDB1 dans le compartiment lymphocytaire T. Avec Bénédicte, j'ai aussi réalisé l'observation initiale d'une plus grande plasticité des lymphocytes Th2 mutants vers le lignage Th1 dans des tests *in vitro* de différenciation et repolarisation.

Lors de ma thèse, le Dr. Binet et moi-même étions en charge de l'aspect cellulaire du projet. Le Dr. Binet a effectué un phénotypage extensif de notre lignée de souris *Setdb1^{fl/fl}xCd4CRE* et a poursuivi les tests de différenciation Th2 et de plasticité. J'ai pour ma part confirmé les observations réalisées *in vitro* en utilisant des chimères mixtes hématopoïétiques comme sources de cellules mutantes et sauvages. Ceci nous a permis d'exclure tout impact de la lymphopénie, observée dans les souris mutante, sur le phénotype des cultures cellulaires. J'ai poursuivi en réalisant des tests de différenciation vers d'autres lignages, ce qui m'a permis de mettre en évidence une exacerbation de la différenciation Th1 dans les cellules *Setdb1^{-/-}*. Enfin, j'ai eu la responsabilité de l'étude de la plasticité *in vivo* du projet, ce qui m'a amené à réaliser des expériences d'immunisation.

Parallèlement à l'aspect cellulaire du projet, le Dr. Adoue était en charge de l'aspect moléculaire. Elle menait des expériences d'immunoprécipitation de la chromatine sur différentes marques épigénétiques dans différents types cellulaires et de transcriptomique pour lesquelles le Dr. Binet et moi-même cultivions les cellules primaires. Bien que l'essentiel des expériences moléculaires ait été effectué par le Dr. Adoue, je me suis formée à ces techniques et j'ai ainsi pu réaliser des expériences de ChIP-qPCR et de RNA-seq. L'analyse bio-informatique des données de séquençages de ChIP-seq et RNA-seq a été coordonnée par le Dr. Joffre et réalisée par le Dr. Adoue et Joanna Fourquet.

Quel que soit le domaine de compétence initial, l'intégralité des membres du groupe a participé activement au développement conceptuel et méthodologique de ce projet. Individuellement nous n'aurions pu mener ce projet à bien et soumettre à publication l'article que je présente dans la partie « Résultats ». Je remercie sincèrement Joanna Fourquet et les Dr. Véronique Adoue, Dr. Bénédicte Binet et Dr. Olivier Joffre pour ces cinq années enrichissantes et ce travail collaboratif.

RESULTATS

L'article présenté ici a été récemment resoumis au journal Immunity après révision.

SETDB1 controls T helper cell lineage integrity by repressing endogenous retroviruses

~~~~~

Véronique Adoue<sup>1,4,5</sup>, Bénédicte Binet<sup>1,4</sup>, Agathe Malbec<sup>1,4</sup>, Joanna Fourquet<sup>1</sup>, Paola Romagnoli<sup>1</sup>, Joost P M van Meerwijk<sup>1</sup>, Sebastian Amigorena<sup>2,3</sup>, Olivier P Joffre<sup>1,5,6</sup>.

<sup>1</sup> CPTP, Université de Toulouse, CNRS, Inserm, UPS, Toulouse, France

<sup>2</sup> Centre de Recherche, Institut Curie, 75005 Paris, France

<sup>3</sup> INSERM, U932, F-75005 Paris, France

<sup>4</sup> These authors contributed equally to this work.

<sup>5</sup> Correspondence: [olivier.joffre@inserm.fr](mailto:olivier.joffre@inserm.fr) (O.J.), [veronique.adoue@inserm.fr](mailto:veronique.adoue@inserm.fr) (V.A.)

<sup>6</sup> Lead contact

Running title: SETDB1 critically controls the Th1 gene expression program

### I. SUMMARY

Upon activation, naïve CD4 T cells differentiate into distinct helper or regulatory T cell subsets depending on environmental signals received. This process relies on complex and lineage-specific gene expression programs, whose dynamics and stability are regulated at the level of the chromatin. The epigenetic pathways involved remain, however, largely unknown. Here, we report that the histone methyltransferase SETDB1 critically controls the Th1 gene expression program. Indeed, SETDB1-deficient naïve CD4 T cells show exacerbated Th1 priming, and when exposed to a Th1-instructive signal, SETDB1-deficient Th2 cells cross lineage boundaries

and transdifferentiate into Th1 cells. Surprisingly, SETDB1 does not appear to directly control Th1 gene promoter activity. Instead, it deposits the repressive H3K9me3 mark at a restricted and cell type-specific set of endogenous retroviruses (ERVs) located in the vicinity of genes involved in immune processes. Refined bioinformatic analyses indicated that these retrotransposons either repress Th1 gene cis-regulatory elements or behave themselves as Th1 gene enhancers. In conclusion, H3K9me3 deposition by SETDB1 ensures T helper cell lineage integrity by repressing a repertoire of ERVs that have been exapted into cis-regulatory modules to shape and control the Th1 gene network.

## II. INTRODUCTION

T lymphocytes protect vertebrates against a wide variety of endogenous and exogenous dangers. Their efficacy comes at least in part from their ability to adapt their phenotype and function to the threat detected by the cells of the innate immune system. Depending on the nature and strength of the signals delivered by these cells and the surrounding tissues, T lymphocytes mobilize different networks of transcription factors, resulting in induction of distinct developmental programs that coordinate the acquisition of lineage-specific and danger-adapted phenotypes and functions (281, 282). This plasticity is best illustrated by naïve CD4 T cells, which are able to differentiate into a large and probably still underestimated number of distinct effector populations.

The transcription factors mobilized in response to environmental signals orchestrate a massive remodeling of the epigenetic landscape of T cells (282, 283). These dynamic changes in chromatin composition and compaction are necessary to set up and stabilize gene expression programs and to allow their faithful transmission to the progeny. Indeed, interfering with the post-translational modifications of histones or with DNA methylation critically affects the differentiation and stability of effector and memory T cells (212, 277, 282, 284). In CD4 T lymphocytes, epigenetic remodeling is largely coordinated by STAT proteins and by the master regulators specific to each lineage, such as T-bet and GATA-3 for, respectively, Th1 and Th2 cells (283, 285). These transcriptional regulators fine-tune the balance between T helper cell determination and plasticity by directing the deposition of permissive epigenetic marks at lineage-specific *cis*-regulatory elements, and by targeting repressive epigenetic pathways to the loci associated with alternative fates (62, 282, 283, 285).

Among the various post-translational modifications found at the N-terminal tail of core histones, the trimethylation of Histone H3 on Lysine 9 (H3K9me3) has emerged as a key

modification that controls genome functions (231). This epigenetic mark was first implicated in the scaffolding and function of constitutive heterochromatin (206, 236). More recently, H3K9me3 deposition at promoters of genes encoding developmental regulators has been described as necessary to repress these loci and maintain embryonic stem cell pluripotency (286). In adult cells, H3K9me3-dependent repression of gene expression in euchromatin and facultative heterochromatin is also important to define and maintain cell identity (212, 287, 288). However, the repertoires of loci and genomic elements that are targeted, as well as the molecular mechanisms at work remain poorly characterized. H3K9me3 finally accumulates on the body of active genes where it may affect transcription elongation and alternative splicing (229, 230). H3K9me3 is thus a versatile chromatin mark, with multiple and sometimes opposed functions.

Several lysine methyltransferases can trimethylate H3K9. They include SUV39H1, SUV39H2 and SETDB1, which all belong to the SUV39H family (231). Whereas SUV39H1 and SUV39H2 were first identified as key components of constitutive heterochromatin (206, 289), SETDB1 was initially found to be involved in the dynamic repression of gene transcription at euchromatin and facultative heterochromatin (211). SUV39H1 has since been shown to repress euchromatic gene expression through H3K9me3 deposition at promoters (212, 287), whereas the maintenance of H3K9me3 at pericentromeric heterochromatin during DNA replication might depend on a stepwise process involving H3K9 mono- and tri-methylation by SETDB1 and SUV39H1, respectively (232). In embryonic stem cells, these two enzymes also collaborate to repress endogenous retroviruses (ERVs) (224). Interestingly, very recent data provided evidence that ERVs have been co-opted as *cis*-regulatory elements to shape and control gene networks in various cell types (290). SETDB1 and SUV39H1 may therefore also control cell integrity through deposition of H3K9me3 at these repeat elements.

The studies mentioned above highlight the general complexity and importance of H3K9me3-dependent epigenetic regulatory pathways in genome function. In T cells, more specifically, SUV39H1 has been implicated in Th2 cell stability by depositing H3K9me3 at the *Ifng* promoter (212). However, the deregulation of the *Ifng* locus observed in *Suv39h1*<sup>-/-</sup> cells cannot by itself explain a loss of Th2 cell integrity. Other critical Th1 cell lineage-specific loci might therefore be controlled by H3K9me3-dependent repressive mechanisms. Interestingly, while a clear H3K9me3 signal is detected at the gene encoding T-bet in Th2 cells, SUV39H1 has no impact in the deposition of the repressive mark at this locus (212). Together with the fact that H3K9me3 disappearance at euchromatin and facultative heterochromatin is limited in SUV39H1-deficient cells (289), these observations suggest that other H3K9me3-dependent epigenetic pathways play a critical role in controlling Th2 cell stability.

Here we set out to investigate the effects of SETDB1-dependent H3K9me3 deposition on CD4 T cell activation, differentiation and commitment. To date, SETDB1 had only been implicated in OX40-dependent repression of the *Ii17a* locus in Th17 cells (277). We show that SETDB1 restricts Th1 cell priming and ensures Th2 cell integrity. Unlike their wild-type counterparts, *Setdb1*<sup>-/-</sup> Th2 cells readily express the Th1 gene network when exposed to the Th1-instructing cytokine IL-12. Interestingly, our data suggest that SETDB1 repress Th1-related loci by depositing H3K9me3 at a subset of ERVs that flank and repress Th1 enhancers or behave themselves as *cis*-regulatory elements of Th1 genes. Therefore, our observations support a model in which H3K9me3 deposition by SETDB1 locks the Th1 gene expression program, and thus ensures T helper cell lineage integrity, by repressing a repertoire of ERVs that have been co-opted to behave as Th1 lineage-specific *cis*-regulatory modules.

### III. RESULTS

#### A. Th1 priming is enhanced in the absence of SETDB1

To analyze the role of SETDB1 in CD4 T cell differentiation and plasticity, we generated mice homozygous for a LoxP-flanked *Setdb1* allele and expressing (*Setdb1*<sup>-/-</sup>), or not (*Setdb1*<sup>+/+</sup>), the CRE recombinase under the control of the *Cd4* promoter. This strategy resulted in the almost complete absence of SETDB1 protein from CD4 single-positive thymocytes (**Figure S1A**). As SETDB1-deficiency was not compensated by overexpression of the other methyltransferases targeting H3K9 (**Figure S1B**), we also observed a marked loss of trimethylated H3K9 from naïve *Setdb1*<sup>-/-</sup> CD4 T cells (**Figure S1C**). The use of the *Cd4-Cre* transgene, which induces SETDB1 deletion relatively late in ontogeny, allowed for normal intrathymic T cell development. Indeed, the total number of cells in the thymus, the relative proportions of the four main populations of thymocytes, and the proportion of mature CD4 T cells were similar in *Setdb1*<sup>-/-</sup> mice and control littermates (**Figures S1D–S1H**).

In peripheral lymphoid tissues, we detected no consequences of T cell-specific SETDB1-deficiency on other populations of immune cells (**Figures S2A–S2C**). SETDB1 was previously implicated in survival in various cell types. For example, conditional deletion of the enzyme in mice expressing the CRE recombinase under the control of the *Mb1* promoter abolished the B cell lineage (291). The impact of *Setdb1* deletion on T cell survival was less pronounced: despite substantially increased activity of caspase-3/7, we only observed a partial loss of the T cell pool (**Figures S2D–S2G**).



To obtain a global view of the changes in gene expression induced by *Setdb1* deletion, we performed RNA sequencing (RNA-seq) on naïve *Setdb1*<sup>-/-</sup> and *Setdb1*<sup>+/+</sup> CD4 T cells. Most of the differentially expressed genes were upregulated in *Setdb1*<sup>-/-</sup> cells (**Figure S3A**), consistent with a globally repressive effect of H3K9me3 on gene transcription. Among the overexpressed genes, those involved in cell division were particularly enriched (**Figures S3B and S3C**). The proportion of CD4 T cells expressing the nuclear antigen Ki67 being higher in *Setdb1*<sup>-/-</sup> mice than in control littermates (**Figure S3D**), this increased expression of cell division-related genes most likely resulted from the observed lymphopenia rather than from a direct impact of *Setdb1* deletion on the regulation of these genes. Moreover, we found no particular enrichment for H3K9me3 domains at cell cycle-related genes in naïve CD4 T cells (**Figure S3E**).

To assess if *Setdb1* deletion could affect T cell function, we next analyzed the differential transcription of a gene set related to T helper cell activation and differentiation. We found no major differences in the expression of these genes between *Setdb1*<sup>-/-</sup> and *Setdb1*<sup>+/+</sup> cells (**Figure 1A and Table S1**), despite the presence of H3K9me3 domains close to loci involved in lymphocyte-mediated immunity (**Figure S3E**). The vast majority of the genes were equally expressed in *Setdb1*<sup>-/-</sup> and *Setdb1*<sup>+/+</sup> cells, most of the differentially expressed loci were transcribed at very low levels, and no lineage-specific transcriptomic signature appeared when focusing on deregulated genes. This lack of effect of SETDB1 deficiency on naïve CD4 T cell programming was further confirmed when we analyzed the production of lineage-specific mediators following acute *ex vivo* stimulation (**Figure 1B**). Together, these observations show that *Setdb1*<sup>-/-</sup> cells are not *a priori* biased towards a specific T helper lineage.

To test whether SETDB1 regulates T helper cell lineage commitment in response to environmental signals, we analyzed *Setdb1*<sup>-/-</sup> and *Setdb1*<sup>+/+</sup> CD4 T cell fate in an IL-12-mediated Th1 differentiation assay. As expected from our experiments measuring caspase 3/7 activity *ex vivo* (**Figures S2F and S2G**), SETDB1 deficiency impaired T cell survival at early time points (**Figure S3F**). However, a significant proportion of cells remained viable and showed normal activation upon TCR triggering (**Figures S3G and S3H**). As T cell differentiation depends on cell cycle progression, we next analyzed the proliferative response of activated CD4 T cells. There were no differences between control and mutant cells (**Figure 1C**), which displayed similar proliferation indexes and percentages of divided cells (**Figure 1D**). Upon exposure to increasing doses of IL-12, *Setdb1*<sup>-/-</sup> cells also expressed T-bet at a similar level as their *Setdb1*<sup>+/+</sup> counterparts (**Figures 1E and 1F**). Together, these observations indicate that SETDB1 deficiency does not affect naïve CD4 T cell activation, proliferation and commitment to the Th1 lineage. It does, however, lead to greater acquisition

of lineage-specific functions. Indeed, both percentages of cells producing Th1-related cytokines and amounts of cytokine synthesized per cell were higher in *Setdb1*<sup>-/-</sup> than in *Setdb1*<sup>+/+</sup> cells (**Figures 1G–1J**). This exacerbated production of cytokines was not the result of a global transcriptional derepression, since we did not evidence any aberrant secretion of soluble mediators related to alternative lineages (**Figure 1J**). It was also not the result of a greater sensitivity of the *Setdb1*<sup>-/-</sup> cells to IL-12 (**Figures S3I–S3K**). Together, these results highlight a key role for SETDB1 in regulating the magnitude of Th1 responses.

## B. Impaired acquisition of the Th2 phenotype by SETDB1-deficient cells

Most of the genes encoding lineage-specific cytokines in naïve CD4 T cells have both permissive and repressive epigenetic marks on their promoters and enhancers. They are thus poised for transcription to guarantee the plasticity, while also preserving the identity of the cells. The enhanced Th1 response observed in *Setdb1*<sup>-/-</sup> cells might, therefore, result from a loss of H3K9me3 at these *cis*-regulatory regions, and this may potentially affect other lineages. To test this hypothesis, we cultured *Setdb1*<sup>+/+</sup> and *Setdb1*<sup>-/-</sup> naïve CD4 T cells in Th2 polarizing conditions. The proliferative response and viability of *Setdb1*<sup>+/+</sup> and *Setdb1*<sup>-/-</sup> cells were comparable at day six (**Figures 2A–2C**). Moreover, *Setdb1*<sup>-/-</sup> cells seemed to commit to the Th2 lineage similar to their control counterparts, with almost all cells expressing GATA-3 (**Figures 2D and 2E**), and no aberrant expression of the Th1-specifying transcription factor T-bet (**Figure 2F**). Production of IL-13 and IL-4 was also similar between *Setdb1*<sup>+/+</sup> and *Setdb1*<sup>-/-</sup> cells (**Figures 2G–2I**). Thus, in contrast to what we observed in Th1 polarizing conditions, there was no enhanced production of Th2 lineage-specific mediators by *Setdb1*<sup>-/-</sup> cells grown in the presence of IL-4. In fact, global transcriptional profiling revealed that the level of expression of the Th2 signature genes was significantly lower in *Setdb1*<sup>-/-</sup> Th2 cells than in their wild-type counterparts, although both wild-type and mutant naïve CD4 T cells efficiently switched on the Th2 program upon exposure to IL-4 (**Figures 2J and 2L**). This impaired induction of the Th2 gene network correlated with lower expression of GATA-3 (**Figures 2D and 2E**) and with decreased chromatin accessibility at Th2 gene enhancers (**Figure 2K**). Interestingly, unlike their wild-type counterparts, *Setdb1*<sup>-/-</sup> cells grown in Th2-polarizing conditions also produced small amounts of the Th1 cytokine IFN- $\gamma$  (**Figures 2G–2I**). This IFN- $\gamma$  ‘leak’ may result from defective repression of Th1-related loci in Th2 cells, which could potentially antagonize the Th2 gene expression program and lead to functional and phenotypic instability.

### C. SETDB1 is required for stable Th2 cell commitment

To assess if SETDB1 controls Th2 cell plasticity, we cultured *Setdb1*<sup>-/-</sup> and *Setdb1*<sup>+/+</sup> cells for six days in Th2-polarizing conditions and then switched to culture in Th1-polarizing conditions. In agreement with the Th1/Th2 paradigm, the control *Setdb1*<sup>+/+</sup> Th2 cells remained stable after two days of culture in Th1 conditions (**Figures 3A–3C**). By contrast, a large fraction of the *Setdb1*<sup>-/-</sup> cell population secreted IFN- $\gamma$ , this phenomenon being even more pronounced after four days of culture (**Figures 3A and 3B**). IFN- $\gamma$  secretion was accompanied by decreased expression of GATA-3 and increased expression of T-bet (**Figures 3D and 3E**). In fact, SETDB1-deficiency allowed the virtually complete reprogramming of Th2 cells upon exposure to Th1-instructing signals, with extinction of Th2 gene expression and induction of a large part of the Th1 gene set (**Figures 3F and 3G and Table S2**). Importantly, this plasticity was not the result of a bias in *Setdb1*<sup>-/-</sup> cells programming due to lymphopenia (**Figures S4A–S4C**).

SETDB1 plays a key role in silencing ERVs (207, 224). Ectopic expression of these retrotransposons can lead to activation of the nucleic acid-sensing machinery and, eventually, to production of type I IFNs (292). Together with IL-12 and IFN- $\gamma$ , type I IFNs can reprogram Th2 cells into stable cells producing IFN- $\gamma$  and expressing both GATA-3 and T-bet (144). Activation of ERVs in *Setdb1*<sup>-/-</sup> Th2 cells might thus account for the increased plasticity that we observed: ERV-induced secretion of IFN- $\alpha$  and IFN- $\beta$  might reprogram the Th2 cells in combination with exogenous IL-12 and the observed aberrant production of IFN- $\gamma$ . However, we found no abnormal levels of type I IFN mRNA in *Setdb1*<sup>-/-</sup> cells (**Figures S4D and S4E**), and neutralization of IFN- $\gamma$  did not prevent *Setdb1*<sup>-/-</sup> Th2 cells from switching to a Th1 phenotype (**Figures S4F and S4G**). To assess more directly if the ectopic expression of Th1-instructive mediators by *Setdb1*<sup>-/-</sup> Th2 cells might account for their phenotypic instability, we co-cultured *Setdb1*<sup>-/-</sup> and *Setdb1*<sup>+/+</sup> Th2 cells in Th1 polarizing conditions. In this setting, the SETDB1-deficient cells still showed substantial plasticity, while their control counterparts did not (**Figures S4H–S4J**). Together, these data indicate that SETDB1 critically controls Th2 cell commitment through a cell-intrinsic mechanism.

To test the role of SETDB1 in CD4 T cell programming *in vivo*, we next immunized *Setdb1*<sup>+/+</sup> and *Setdb1*<sup>-/-</sup> mice with the 1W1K variant of the I-E alpha chain immunodominant peptide formulated in RIBI adjuvant. Using 1W1K/I-A<sup>b</sup> tetramers, we first showed that antigen-specific *Setdb1*<sup>-/-</sup> CD4 T cells expanded and accumulated in the lymph nodes draining the site of immunization (**Figures 3H and 3I**). The frequency of tetramer-positive cells was lower than

in control mice but this quantitative defect did not prevent SETDB1-deficient cells from differentiating into Th1 and Th2 effector cells. However, relative to the size of the antigen-specific T cell compartment, the frequency of CD4 T cells producing IFN- $\gamma$  was strongly increased in *Setdb1*<sup>-/-</sup> as compared to *Setdb1*<sup>+/+</sup> mice (**Figure 3J**). This exacerbated Th1 response shifted the Th1/Th2 balance towards Th1 dominant immunity (**Figure 3K**). Importantly, lymphopenia is very unlikely to explain this effect as CD4 T cell responses are skewed toward the Th2 phenotype in mice and humans with reduced number of peripheral T cells (293). The exacerbated Th1 response observed in *Setdb1*<sup>-/-</sup> mice is also unlikely to result from an over-representation of recent thymic emigrants within the peripheral T cell pool, since these cells are also biased toward the Th2 effector lineage (25). Therefore, as observed *in vitro*, SETDB1 controls the Th1 gene expression program *in vivo* and thus regulates the equilibrium between Th1 and Th2 immune responses.

#### D. SETDB1-dependent H3K9 trimethylation at a subset of ERVs

To determine how SETDB1 controls Th2 cell commitment and stability, we performed high-throughput chromatin immunoprecipitation (ChIP) of H3K9me3 in *Setdb1*<sup>+/+</sup> and *Setdb1*<sup>-/-</sup> Th2 cells. As H3K9me3 deposition at the *Ifng* promoter has been linked to Th2 cell stability (212), we first analyzed H3K9me3 signal at gene promoters. As expected from the literature, our genome-wide analysis revealed an inverse correlation between the deposition of the mark and gene expression (**Figure S5A**). However, SETDB1-deficiency did not significantly impact on H3K9me3 deposition at Th1-related genes (**Figure S5B**). We then analyzed the genome-wide distribution of H3K9me3 domains in *Setdb1*<sup>+/+</sup> Th2 cells. We observed no more peaks at gene bodies or promoters than would be expected by a random distribution (**Figures 4A and 4B**). By contrast, we found statistically significant enrichment of H3K9me3 domains at enhancers (defined as non-promoter H3K4me1<sup>+</sup> genomic regions) and at ERVs (**Figures 4A and 4B and Table S3**). Interestingly, 74% of the peaks at enhancers also overlapped with ERVs (**Figure 4C**), suggesting that the ERVs rather than the enhancers themselves may be the main targets for H3K9 trimethylation. To test this hypothesis, we analyzed the distribution of H3K9me3 across the length of individual H3K9me3<sup>+</sup> ERV and enhancer sequences. The H3K9me3 signal clearly peaked at and aligned with the center of the ERVs (**Figures 4D and 4E**). By contrast, the signal appeared randomly distributed across the enhancer sequences, with only a modest accumulation on their flanking regions. These data strongly suggest that H3K9 trimethylation is directed at ERVs and marks enhancers only when they overlap or flank the retrotransposons. Interestingly, in agreement with the repressive role of the histone mark and with the status of the *cis*-regulatory elements, H3K9me3 signal was higher on the flanks of poised than of active

enhancers, and the mark only overlapped inactive *cis*-regulatory elements (**Figure 4F**). H3K9me3 accumulation over the enhancer sequence might thus result from a spreading from neighboring ERVs (**Figure 4G**), or may be due to physical overlap between the enhancers and the retroelements (**Figure 4H**). Whatever the model, we observed that H3K9me3 domains were closer to the center of the ERVs than to the center of the enhancers (**Figure 4I**). Therefore, these results support the hypothesis that H3K9me3 is deposited at a subset of ERVs that overlap or flank enhancers, which might lead to their repression.

Finally, to determine which lysine methyltransferase is necessary for H3K9me3 deposition at ERVs, we determined if the retroelements marked by H3K9me3 in wild-type cells were still covered by the repressive mark in *Setdb1*<sup>-/-</sup> or *Suv39h1*<sup>-/-</sup> Th2 cells. Interestingly, SUV39H1-deficiency had no major impact on H3K9me3 deposition at these genomic locations. In contrast, most of the peaks disappeared from the ERVs in the SETDB1-deficient cells (**Figure 4J**), even if some residual signal could persist at certain locations (**Figures S6A–S6C**). Together, these data indicate that SETDB1 targets H3K9me3 at a subset of ERVs in Th2 cells, and that some of these retrotransposons overlap or flank enhancers

## E. SETDB1-deficiency upregulates ERVs and their neighboring genes

Recent studies have provided evidence for the hypothesis that transposable elements have been co-opted for the regulation of host gene networks (290, 294). The impact of SETDB1 deletion on CD4 T cell fate might, therefore, result from a loss of H3K9me3 at ERVs that behave as *cis*-regulatory modules of transcription, and/or regulate the activity of enhancers. To test this hypothesis, we analyzed the consequences of SETDB1 deletion on ERV accessibility and activity, on the status of their nearest enhancers, and on the expression of associated genes. We first compared ERV expression levels in *Setdb1*<sup>-/-</sup> and *Setdb1*<sup>+/+</sup> Th2 cells. In *Setdb1*<sup>-/-</sup> Th2 cells, the expression of 22% of the ERVs that lost H3K9me3 was deregulated, 77% of which being overexpressed (**Figures 5A and 5B**). As the expression levels were very low, we used other parameters to confirm that H3K9me3 disappearance from ERVs lead to local chromatin remodeling. Loss of H3K9me3 at ERVs in *Setdb1*<sup>-/-</sup> Th2 cells also correlated with the accumulation and spreading of permissive histone marks, including H3K4me1 and H3K27ac, and with increased chromatin accessibility (**Figures 5C and S6A–S6C**). Interestingly, we also observed a decompaction of the chromatin over several kilobases on both sides of the retrotransposons overexpressed in mutant cells (**Figure 5C**). We therefore hypothesized that the *cis*-regulatory elements that flank ERVs might be de-repressed

in the absence of SETDB1. To test this hypothesis, we analyzed bidirectional expression of enhancers located in the vicinity of ERVs marked by H3K9me3 in *Setdb1*<sup>+/+</sup> cells and upregulated in *Setdb1*<sup>-/-</sup> cells. The enhancers associated with ERVs that were activated following H3K9me3 disappearance were themselves more expressed in *Setdb1*<sup>-/-</sup> than in *Setdb1*<sup>+/+</sup> Th2 cells (**Figure 5D**). As expected from their increased transcription, they also accumulated permissive histone marks in mutant cells (**Figure 5E**). Together these data show that SETDB1 deletion leads to loss of H3K9me3 from ERVs, and to the concomitant activation of these retroelements and of their neighboring enhancers. Finally, we tested if this cascade of events resulted in deregulation of gene expression. Even if the Th2 gene expression program is globally down-regulated in *Setdb1*<sup>-/-</sup> Th2 cells (**Figures 2J and 2L**), our analysis showed that the genes associated with enhancers flanking or overlapping ERVs overexpressed in SETDB1-deficient cells were also significantly more expressed (**Figure 5F**). Finally, supporting the hypothesis that ERVs behave as *cis*-regulatory modules controlling gene expression, we observed a positive correlation between the degree of gene expression change per gene between *Setdb1*<sup>-/-</sup> and *Setdb1*<sup>+/+</sup> cells and the number of activated retroelements associated with a gene (**Figure 5G**). We propose that SETDB1-dependent H3K9me3 deposition at ERVs inactivates neighboring enhancers and thus participates to the silencing of their target genes.

## F. H3K9me3+ ERVs mark the Th1 enhancer landscape in Th2 cells.

In our functional assays, SETDB1 deletion led to enhanced Th1 priming and to Th2 cell instability. As discussed above, loss of regulation of the Th1 gene network may underlie these observations. Based on our epigenetic and transcriptomic studies, we postulated that SETDB1 controls Th1 gene expression by either repressing ERVs operating as *cis*-regulatory elements of these genes, and/or regulating the activity of their enhancers. To test this hypothesis, we first assigned biological significance to the ERVs marked by H3K9me3 in a SETDB1-dependent manner by analyzing the annotations of their nearby genes. We observed a strong association of the retrotransposons with genes involved in immune processes, including leukocyte activation and cytokine production (**Figure 6A**). Interestingly, this distribution was cell type-specific: there was very little overlap between the ERVs marked by H3K9me3 in Th2 cells and those marked in adipocytes (**Figure S7A**). Moreover, the ERVs marked by H3K9me3 in white adipose cells were associated with genes that have no direct link with immunity (**Figures S7B and S7C**). In Th2 cells, thus, SETDB1 is responsible for H3K9 trimethylation at a restricted and cell type-specific set of ERVs that are associated with genes controlling T cell functions. As IFN- $\gamma$  plays a critical role in Th1 cell programming, we next investigated whether these ERVs more



specifically targeted IFN- $\gamma$ -stimulated genes (ISGs). In contrast to H3K9me3<sup>-</sup> ERVs or to the repertoire of H3K9me3<sup>+</sup> ERVs found in adipocytes, the ERVs marked by the repressive histone mark in Th2 cells were located in the vicinity of ISGs (**Figure 6B**). Motif enrichment analysis of H3K9me3-marked ERV sequences strengthened this observation; it revealed a strong enrichment for the binding sites of STAT1, the main transcription factor responsible for the diverse cellular effects induced by IFN- $\gamma$  (**Figure 6C**), as well as for other critical Th1-related transcription factors, including T-bet, FOXO3, IRF1 and STAT4. The unbiased “Upstream Regulator Analysis” of our RNA-seq data also identified these five transcription factors as very likely to be responsible for the differences in gene expression observed in *Setdb1*<sup>-/-</sup> vs *Setdb1*<sup>+/+</sup> Th2 cells upon culture in Th1-polarizing conditions (**Figures 6D** and **S7D**). To strengthen these *in silico*-based predictions, we next tested whether ERV sequences marked by H3K9me3 in Th2 cells were associated with the genomic localizations of STAT1 and STAT4 in Th1 cells. In contrast to H3K9me3<sup>-</sup> ERVs or to the repertoire of H3K9me3<sup>+</sup> ERVs found in adipocytes, we observed that the ERVs marked by the repressive histone mark in Th2 cells were strongly enriched near Th1-specific STAT1 and STAT4 binding sites (**Figures S7E** and **S7F**). In fact, almost 2,000 H3K9me3<sup>+</sup> retroelements are associated with Th1-specific STAT1/4 genomic locations in Th2 cells (**Figures 6E–6H**). Interestingly, whereas a substantial number of these ERVs overlapped STAT ChIP-seq peaks, and were thus very likely to behave as Th1 gene enhancers, most of them only flanked the transcription factors binding sites (**Figure S7G**). Therefore, ERVs probably mainly regulate the Th1 gene network by modulating the accessibility and activity of the Th1 enhancers located in their vicinity. To definitely confirm that SETDB1 deposits H3K9me3 at a subset of ERVs associated with Th1 enhancers in Th2 cells, we analyzed the location of the retroelements relative to Th1 enhancers. We found 4,411 putative Th1 enhancers associated with H3K9me3<sup>+</sup> ERVs in Th2 cells (**Figure 6I**). In contrast, we did not detect any enrichment of the Th1-specific *cis*-regulatory elements at H3K9me3<sup>+</sup> ERVs in adipocytes (**Figure 6J**). Together these data suggest that the differences in stability observed between *Setdb1*<sup>-/-</sup> and *Setdb1*<sup>+/+</sup> Th2 cells are very likely explained by SETDB1 causing H3K9me3 deposition at, and thus suppression of, ERVs that overlap or flank a large network of Th1-specific enhancers. This epigenetic silencing pathway is probably central for the H3K9me3-dependent suppression of the Th1 gene network in Th2 cells, since 70% of the Th1 *cis*-regulatory elements that are covered by the repressive mark overlap an ERV onto which H3K9me3 is targeted (**Figures 6K** and **6L**).

## G. SETDB1-dependent H3K9 trimethylation at ERVs represses Th1-specific enhancers.

H3K9me3<sup>+</sup> ERVs mark a large repertoire of Th1 enhancers in Th2 cells. To test whether the deposition of H3K9me3 at these genomic elements is associated with their repression, we first analyzed the activation status of the Th1 enhancers associated with H3K9me3<sup>+</sup> ERVs in wild-type Th2 cells. As expected, most of these *cis*-regulatory elements were repressed (**Figure 7A**), and we detected an accumulation of the histone mark over “poised” or “ghost” enhancers but not on active *cis*-regulatory elements (**Figure 7B**). To test whether the histone mark had a causal role in the repression of Th1 enhancers, we next compared their status in *Setdb1*<sup>+/+</sup> and *Setdb1*<sup>-/-</sup> Th2 cells. Interestingly, we observed that H3K9me3 disappearance in mutant cells led to a strong increase in chromatin accessibility at Th1 enhancers associated with ERVs that were upregulated in a mutant context (**Figure 7C**). This derepression correlated with an accumulation of the permissive histone marks H3K4me1 and H3K27ac on the enhancer sequences (**Figures 7D–G**). From these data, we conclude that the SETDB1-dependent deposition of H3K9me3 at ERVs in Th2 cells is necessary to repress a large network of Th1 enhancers, and thus, to guarantee lineage commitment and stability. These enhancers included the previously identified conserved non-coding sequence (CNS) located 17–20kb downstream of the *Ifng* gene (CNS<sub>17-20</sub>). While this region is poised in naïve T cells, with a strong H3K4me1 signal flanked by a large domain of H3K9me3, it loses competence upon Th2 cell commitment, with an accumulation of H3K9me3 and a complete loss of H3K4me1 (**Figure 7F**). In *Setdb1*<sup>-/-</sup> cells, this enhancer region as well as the hundreds of others marked or flanked by an ERV, lose H3K9me3 and thus probably become accessible to the Th1-specific transcription factors mobilized downstream of the IFN-γ and IL-12 receptor. Consistent with this hypothesis, we found that the H3K4me1 and ATAC-seq signals at *Ifng* CNS<sub>17-20</sub> was substantially higher in *Setdb1*<sup>-/-</sup> than in *Setdb1*<sup>+/+</sup> Th2 cells (**Figures 7G and 7H**). It therefore appears that the lack of deposition of H3K9me3 at ERVs overlapping or flanking Th1-related enhancers in the absence of SETDB1 makes the latter accessible to transcription factors, including IRF1, T-bet, FOXO3 and lineage-associated STATs. This lack of repression potentially affects a large number of Th1-associated loci, including those encoding IFN-γ and T-bet, signature cytokine and “master regulator” of the lineage, respectively. They also include those encoding other critical transcriptional regulators such as STAT4, IRF1 and RUNX3 (**Figure 7I**). These genes have at least one enhancer associated with an ERV marked by H3K9me3 in Th2 cells and are, in addition, more expressed in *Setdb1*<sup>-/-</sup> than in *Setdb1*<sup>+/+</sup> Th2 cells upon culture in Th1-inducing conditions. In conclusion, our data reveal that Th2 cell stability is controlled at



the level of chromatin by the SETDB1-dependent deposition of H3K9me3 at a restricted set of ERVs flanking or behaving as Th1-gene enhancers.

## IV. DISCUSSION

Up to 10% of the mouse genome is comprised of ERVs, which have long been considered to be junk DNA sequences. Recently, however, regulatory functions over gene expression have been assigned to these transposable elements. In mouse CD4 T cells, we document for the first time that a set of ERVs enriched in binding sites for pro-Th1 transcription factors overlap or flank the enhancers of genes from the Th1 cell transcriptomic signature. We further show that the accessibility of these repeat elements is regulated at the epigenetic level by SETDB1. Indeed, in SETDB1-deficient cells, the lack of deposition of the repressive mark H3K9me3 at this subset of ERVs correlates with their activation and with the overexpression of their closest genes. At the cellular level, this deregulation of gene expression due to SETDB1-deficiency translates into increased Th2 cell plasticity and enhanced Th1 cell priming. Together, these data suggest that SETDB1 controls Th2 cell integrity by repressing a restricted and cell-type specific repertoire of ERVs.

ERVs potentially control Th1 genes expression through two non-mutually exclusive mechanisms: they may either behave as *cis*-regulatory elements, or they may regulate chromatin accessibility at nearby enhancers. Our evidence that the binding motifs for critical Th1-associated transcription factors are enriched in H3K9me3<sup>+</sup> ERV sequences supports the idea that the ERVs directly act as *cis*-regulatory elements. The existence of such a subset of regulatory ERVs, which might have shaped the Th1 transcriptional network over time, is supported by a recent study showing that ERVs containing binding sites for IFN-induced transcription factors are necessary for AIM2 inflammasome activation (294). However, when we fractionated the repertoire of H3K9me3<sup>+</sup> ERVs associated with Th1 enhancers in Th2 cells, we observed that most of them only flanked STAT-1/4 binding sites. This result suggests that ERVs may regulate the Th1 gene network mainly by modulating the activity of the Th1 enhancers located in their vicinity. Interestingly, the H3K9me3<sup>+</sup> ERVs that flank Th1 enhancers seem to accumulate at a distance of 3 to 5 kb from the STAT peaks. This distribution of the retrotransposons fully overlaps with the distribution of the H3K9me3 signal observed on the flanking regions of enhancers whose activity is regulated by this histone mark in DC and fibroblasts (216). Although the authors of that study did not implicate SETDB1 in H3K9me3 deposition and did not identify ERVs as the targeted genomic elements, they did correlate the accumulation of H3K9me3 at this location with the repression of adjacent enhancer activity.

This observation reinforces, and extends to other cell types, our model supporting that retrotransposons are the genetic elements that are targeted by the H3K9me3-dependent silencing machinery to regulate enhancer activity in a cell type-specific manner. The underlying molecular mechanism probably relies on local heterochromatin spreading from ERVs to nearby regulatory elements as suggested by our ChIP-seq data and by studies from the literature (222).

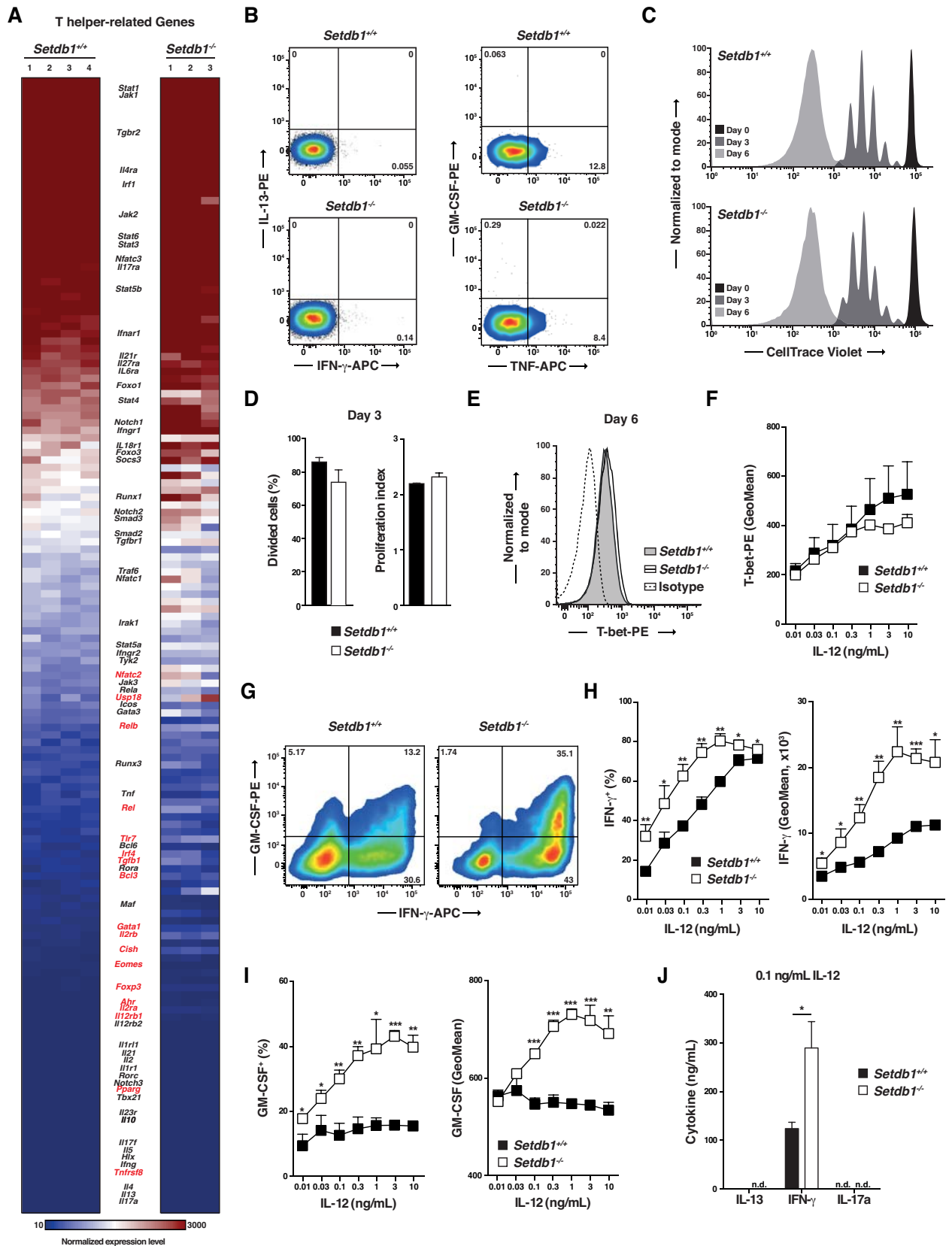
SETDB1 is ubiquitously expressed in mouse tissues and more than 900,000 ERVs are dispersed through the mouse genome. Nevertheless, the repertoire of ERVs under the control of SETDB1 is limited and highly cell type-specific. Only 17,349 ERVs, or 1.9% of all those in the genome, are associated with H3K9me3 domains in Th2 cells. The vast majority of these ERVs are not associated with H3K9me3 domains in white adipose cells. Moreover, the biological functions of genes associated with the ERVs marked by H3K9me3 in Th2 and white adipose cells are fully different: they are associated with genes involved in immune processes in differentiated lymphocytes, whereas they have no direct link with immunity in adipocytes. The enrichment of H3K9me3 at a specific set of ERVs might be explained by the fact that SETDB1 is recruited to the chromatin by Krüppel-associated box zinc-finger proteins (KZFPs) using the scaffold protein TRIM28 as a molecular intermediate. The mouse genome encodes hundreds of KZFPs, whose expression depends on the cell type and its physiological state (295). These transcriptional regulators have different DNA binding sites and ERVs are one of their main genomic targets. Indeed, KZFPs have co-evolved with transposable elements and use evolutionarily conserved regions located mainly within their regulatory sequences to control gene expression (290, 295). The cell-type specific and SETDB1-dependent deposition of H3K9me3 at ERVs that we observed in Th2 cells is therefore probably orchestrated by a specific set of KZFPs that await identification.

Of the three lysine methyltransferases from the SUV39H family, our data show that only SETDB1 is necessary for silencing ERVs in differentiated lymphocytes. These findings are consistent with those obtained in neural progenitor cells and immortalized mouse embryonic fibroblasts, in which SUV39H1-deficiency did not severely affect ERV silencing (224). Although they apparently argue against a direct collaboration of SUV39H1 and SETDB1 in H3K9 trimethylation at ERVs, our data do not fully exclude cooperation between these two enzymes in regulating Th2 cell commitment. In fact, the H3K9me3-dependent epigenetic regulation of CD4 T cell differentiation involves both SUV39H1 and SETDB1. In differentiated Th2 cells, it has been shown that SUV39H1 controls H3K9me3 deposition at the *Ifng* promoter (212) and our data demonstrate that SETDB1 regulates the entire Th1 gene network through repression of ERVs overlapping or flanking Th1-specific enhancers. To guarantee Th2 cell stability in a

changing environment, two non-redundant epigenetic silencing pathways therefore converge to lock the Th1 transcriptional program at different genomic locations.

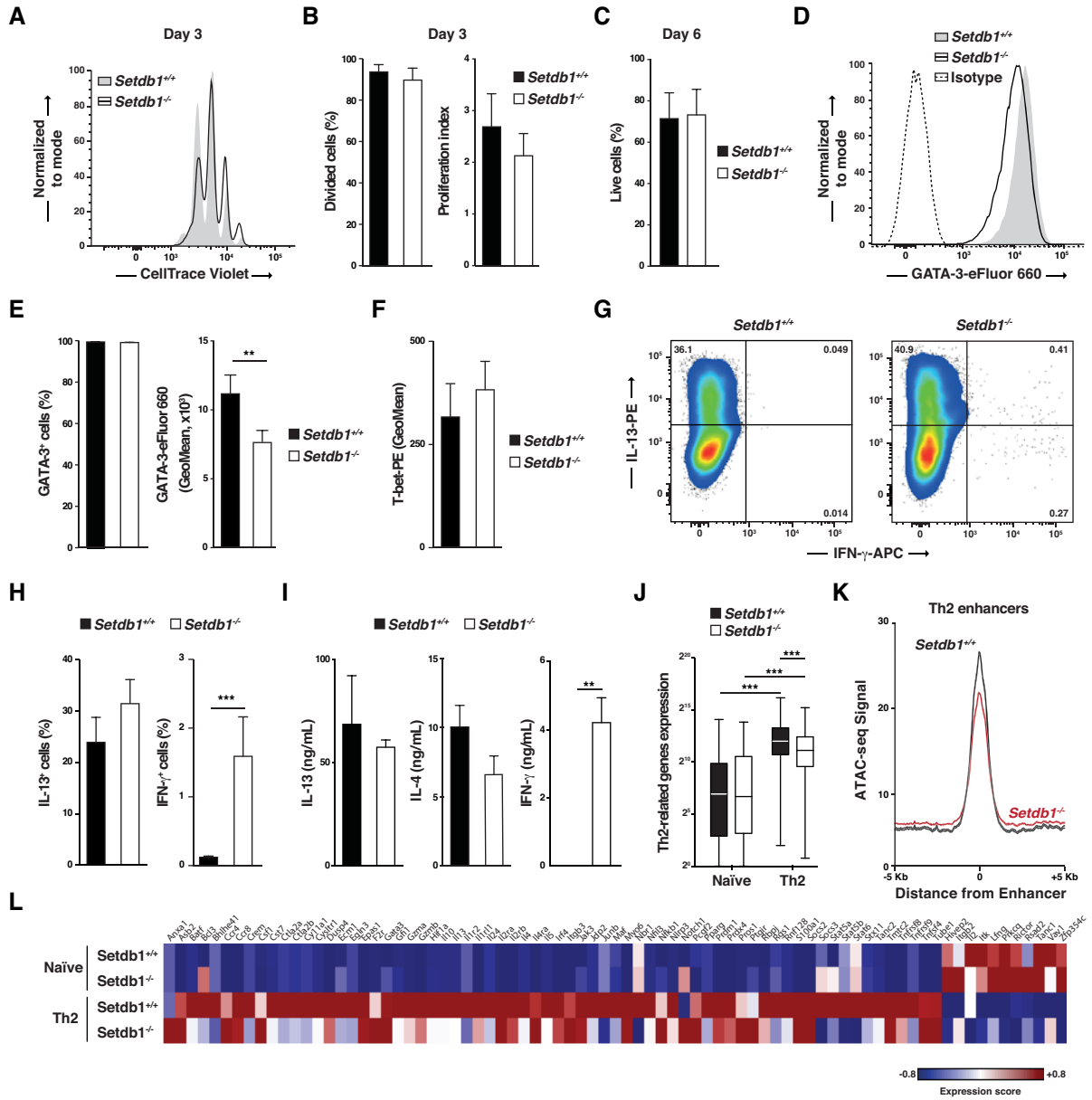
In conclusion, we have shown here that SETDB1 controls CD4 T cell identity by repressing ERVs that flank or overlap Th1 lineage-specific enhancers. This enzyme is thus a potential target for drugs that might be useful, for example, to promote Th1 cell differentiation in various infectious diseases, or to prevent harmful Th2 responses in allergic disorders.

## V. FIGURES



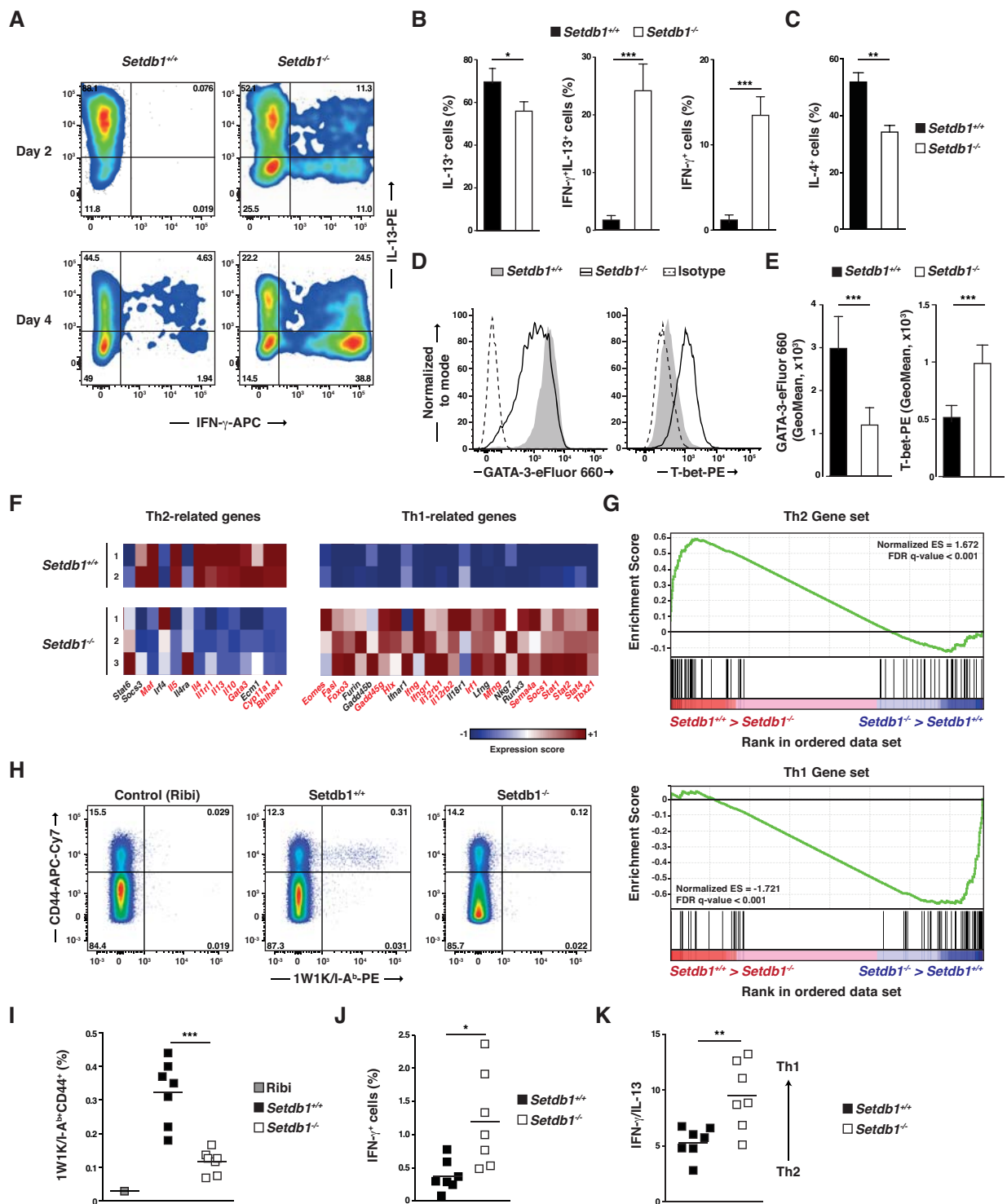
**Figure 1. Th1 priming is enhanced in the absence of SETDB1.**

(A) Expression of T helper-related genes by *Setdb1*<sup>+/+</sup> and *Setdb1*<sup>-/-</sup> naïve CD4 T cells as measured by RNA-seq. The names of the genes that are differentially expressed in the two genotypes (adjusted p-value<0.1 and fold difference >2 or <0.5) are highlighted in red. (B) Cytokine production by *Setdb1*<sup>+/+</sup> and *Setdb1*<sup>-/-</sup> naïve CD4 T cells, as measured by intracellular staining and flow cytometry following acute ex vivo stimulation. (C) Proliferation profiles of *Setdb1*<sup>+/+</sup> and *Setdb1*<sup>-/-</sup> CD4 T cells cultured for 0, 3 or 6 days in Th1 medium. (D) Percentage of divided cells (left) and proliferation index (right) as calculated from the data in (C). (E) Expression of T-bet by *Setdb1*<sup>+/+</sup> and *Setdb1*<sup>-/-</sup> CD4 T cells after 6 days of culture in Th1 medium (0.3 ng/mL IL-12), as measured by intracellular staining and flow cytometry. (F) Average expression of T-bet by *Setdb1*<sup>+/+</sup> and *Setdb1*<sup>-/-</sup> CD4 T cells after 6 days of culture in Th1 medium containing increasing concentrations of IL-12. (G) Production of GM-CSF and IFN- $\gamma$  by *Setdb1*<sup>+/+</sup> and *Setdb1*<sup>-/-</sup> CD4 T cells after 6 days of culture in Th1-inducing conditions (0.3 ng/mL IL-12). Cytokine production was measured by intracellular staining following acute restimulation. (H, I) Percentage of CD4 T cells-producing cytokine (left) and average cytokine production per cell (right) after 6 days of culture in Th1 medium containing increasing concentrations of IL-12. (J) Production of cytokines by *Setdb1*<sup>+/+</sup> and *Setdb1*<sup>-/-</sup> CD4 T cells following 6 days of culture in Th1-inducing conditions and overnight restimulation with anti-CD3e and anti-CD28 antibodies, as measured in cell culture supernatants. Data are representative of two (C) or three (B, E and G) independent experiments. Data are represented as mean  $\pm$  SD of two (D) or three (J) independent experiments or of three biological replicates from one representative experiment out of three performed (F, H, I). \* $p$ <0.05, \*\* $p$ <0.01, \*\*\* $p$ <0.001 (unpaired Student's t-test). See also Figure S1-S3 and Table S1.



**Figure 2. Impaired acquisition of the Th2 phenotype by SETDB1-deficient cells.**

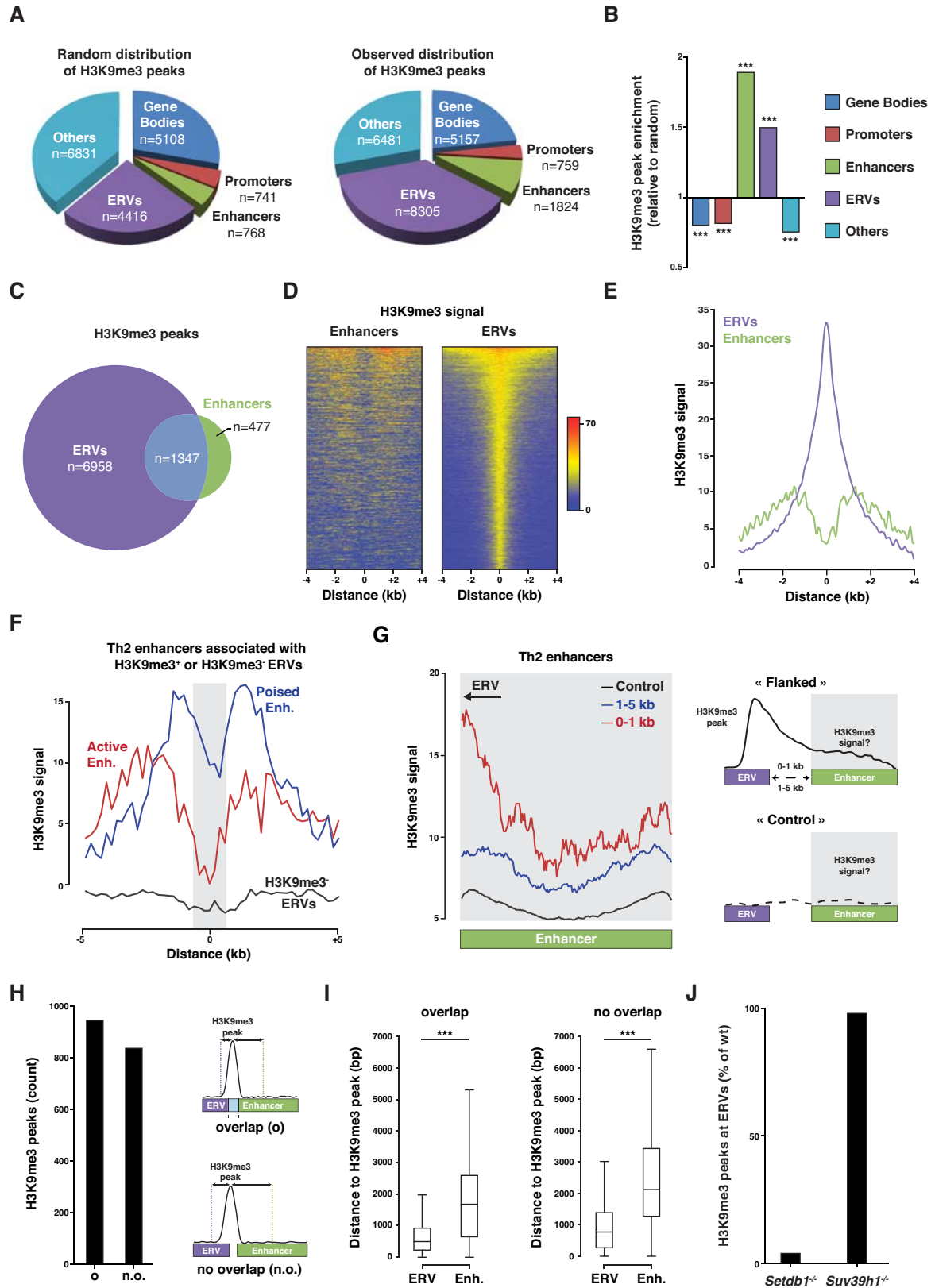
*Setdb1*<sup>+/+</sup> and *Setdb1*<sup>-/-</sup> naïve CD4 T cells were cultured for 3 (A, B) or 6 (C–H) days in Th2-polarizing conditions. (A) Proliferation profiles, as determined by CTV dilution. (B) Percentages of divided cells and proliferation index, as calculated from the data in (A). (C) Percentage of live (acridin orange-positive and propidium iodide-negative) cells. (D) Expression of GATA-3, as determined by intracellular immunostaining and flow cytometry. (E) Percentage of GATA-3-expressing cells and average expression of GATA-3 per cell. (F) Average expression of T-bet, as measured by intracellular immunostaining and flow cytometry. (G) Production of IL-13 and IFN- $\gamma$  by *Setdb1*<sup>+/+</sup> and *Setdb1*<sup>-/-</sup> CD4 T cells after 6 days of culture in Th2-inducing conditions, as determined by intracellular staining and flow cytometry following acute restimulation. (H) Percentage of cytokine-producing CD4 T cells. (I) Production of Th1- (IFN- $\gamma$ ) and Th2-related (IL-13, IL-4) cytokines by *Setdb1*<sup>+/+</sup> and *Setdb1*<sup>-/-</sup> CD4 T cells following 6 days of culture in Th2-polarizing conditions and overnight restimulation with anti-CD3e and anti-CD28 antibodies, as measured in cell culture supernatants. (J) Box and whisker (min to max) diagrams representing the mRNA levels of 79 Th2-related genes in *Setdb1*<sup>+/+</sup> and *Setdb1*<sup>-/-</sup> naïve and Th2 cells. (K) Average ATAC-seq signal at Th2 enhancers (defined as H3K4me1<sup>+</sup> regions in Th2 cells that do not overlap with gene promoters) in *Setdb1*<sup>+/+</sup> and *Setdb1*<sup>-/-</sup> Th2 cells. (L) Th2-related genes expression in *Setdb1*<sup>+/+</sup> and *Setdb1*<sup>-/-</sup> naïve and Th2 cells. Data are representative of three (A) or eight (D, G) independent experiments. Data are represented as mean  $\pm$  SEM of eight independent experiments (E, F and H) or are means  $\pm$  SD from three (I), five (B) or six (C) independent biological replicates from three independent experiments. \*\* $p$ <0.01, \*\*\* $p$ <0.001 (paired Student's t-test). The data in (J, L) and (K) are from three and two independent biological replicates, respectively.





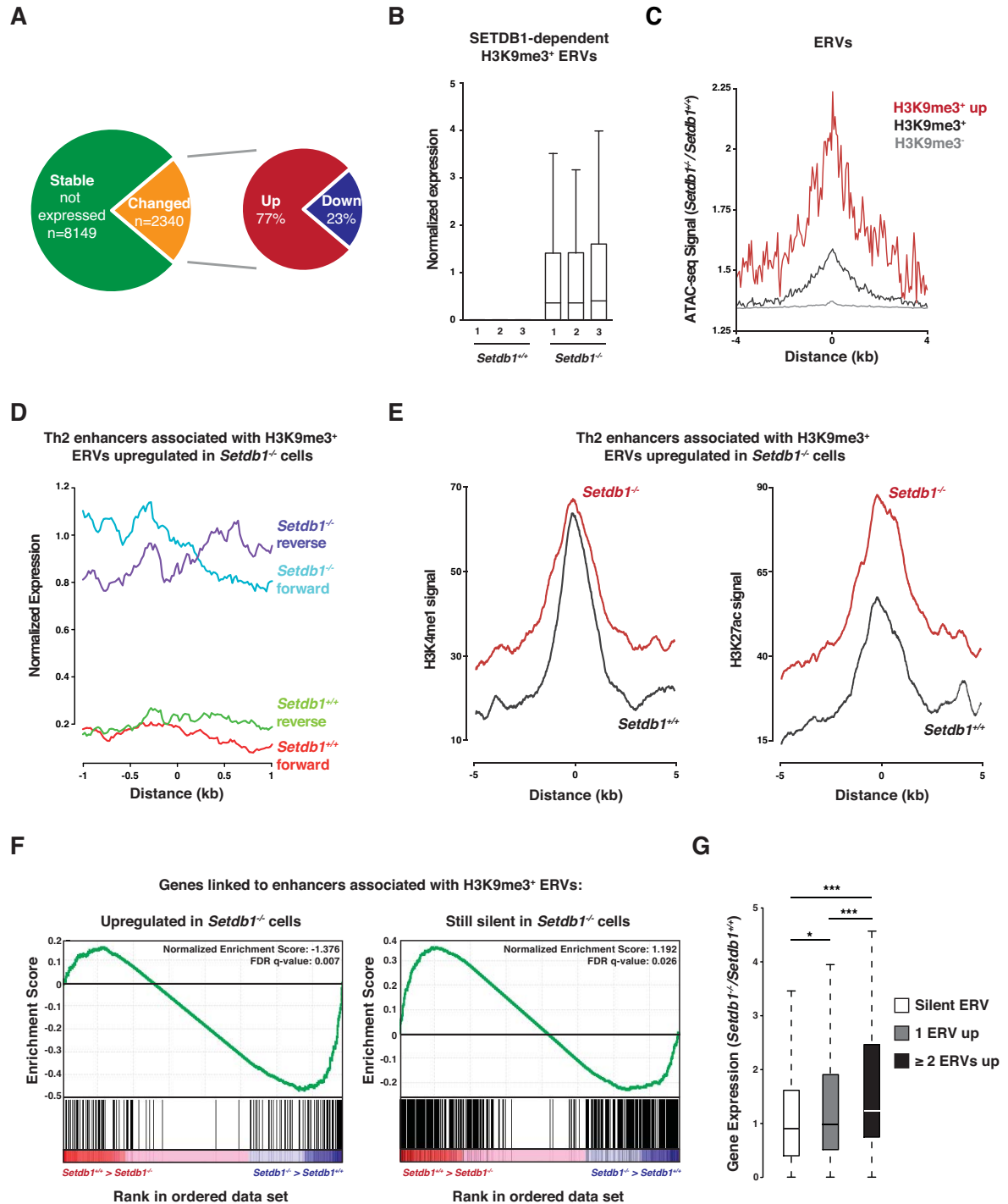
**Figure 3. SETDB1 deficiency leads to impaired Th2 cell commitment *in vitro* and to deregulation of the Th1/Th2 balance *in vivo*.**

(A-G) Unless stated otherwise, *Setdb1*<sup>+/+</sup> and *Setdb1*<sup>-/-</sup> Th2 cells were analyzed after being cultured for two days in Th1 polarizing conditions. (A) Production of IL-13 and IFN- $\gamma$ , as determined by intracellular immunostaining and flow cytometry after two days (top) and four days (bottom) of culture in Th1 medium. (B) Percentages of cells producing IL-13 and/or IFN- $\gamma$  after two days of culture, as calculated from the data in (A). (C) Percentages of cells producing IL-4, as determined by intracellular immunostaining and flow cytometry. (D) Expression of GATA-3 and T-bet, as determined by intracellular immunostaining and flow cytometry. (E) Average expression of GATA-3 and T-bet, as calculated from the data in (D). (F) Th2- and Th1-related gene expression. Differentially expressed genes (adjusted p-value <0.1 and fold difference >2 or <0.5) are indicated by the gene names in red. (G) GSEA of *Setdb1*<sup>-/-</sup> Th2 cells, with genes ranked on the basis of expression in *Setdb1*<sup>-/-</sup> Th2 cells relative to that in their *Setdb1*<sup>+/+</sup> counterparts. (H-K) *Setdb1*<sup>+/+</sup> and *Setdb1*<sup>-/-</sup> mice were immunized s.c. with 1W1K peptide in RIBI adjuvant. Eight days after, the lymph nodes draining the site of immunization were collected and the antigen-specific CD4 T cell response was analyzed. The frequency of antigen-specific cells among CD4 T cells was first determined by flow cytometry using the 1W1K/I-A<sup>b</sup> tetramer. Representative dot-plots (H) and frequency of 1W1K/I-A<sup>b</sup>CD44<sup>+</sup> cells among CD4 T cells for each mouse (I) are shown. (J) Percentages of IFN- $\gamma$  producing cells among 1W1K/I-A<sup>b</sup> CD4 T cells upon antigen-specific restimulation, as determined by intracellular immunostaining and flow cytometry. The percentages are expressed relative to the size of the antigen-specific CD4 T cell compartment, as determined in (I). (K) IFN- $\gamma$  and IL-13 production by CD4 T cells was determined by ELISA following *ex vivo* antigen-specific restimulation. The IFN- $\gamma$  to IL-13 ratio is shown for each mouse. Data are representative of eight independent experiments (A, D) or are represented as mean  $\pm$  SEM (B, E) of eight independent experiments or as mean  $\pm$  SD of three independent biological replicates (C). The data in (F) and (G) are from two (*Setdb1*<sup>+/+</sup>) or three (*Setdb1*<sup>-/-</sup>) independent biological replicates. \**p*<0.05, \*\*\**p*<0.001 (unpaired Student's t-test). See also Figure S4 and Table S2.



**Figure 4. SETDB1-dependent H3K9 trimethylation at a subset of ERVs.**

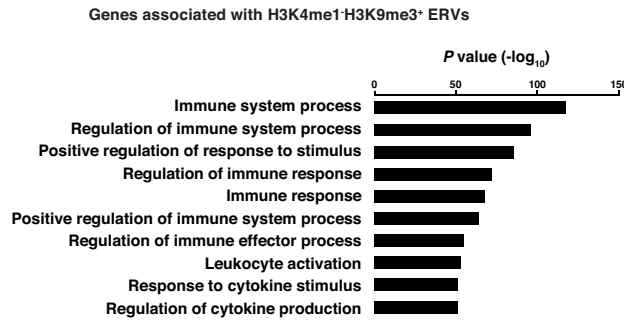
(A) Random and observed genomic distribution of H3K9me3 domains in *Setdb1*<sup>+/+</sup> Th2 cells, as measured by ChIP-seq. (B) Relative enrichment of H3K9me3 domains at indicated genomic elements. \*\*\* $p < 0.001$  Pearson's chi-squared test. (C) Numbers of H3K9me3 domains found at ERVs, enhancers, or both. (D) H3K9me3 signal distribution across each H3K9me3<sup>+</sup> enhancer or ERV sequence. The "0" on the X-axis corresponds to the center of the indicated genomic element. (E) Average H3K9me3 signal profiles at H3K9me3<sup>+</sup> ERVs and enhancers. (F) Average H3K9me3 signal profiles at Th2 enhancers associated with H3K9me3<sup>-</sup> ("H3K9me3<sup>-</sup> ERVs") or H3K9me3<sup>+</sup> ERVs showing ("Active Enh.") or not ("Poised Enh.") enrichment for H3K27ac in Th2 cells. The "0" on the X-axis corresponds to the center of the enhancer. (G) H3K9me3 signal (median) at enhancers that do not overlap with H3K9me3 domains and that are located at a distance of 0 to 1kb (red), 1 to 5kb (blue) or more than 5kb (black) from an ERV mark by H3K9me3. (H) Number of H3K9me3 domains overlapping ERVs and enhancers which themselves overlapped (o) or are mutually exclusive (non-overlapping, n.o.). (I) Box plots representing the distance between H3K9me3 domains and ERVs or enhancers in the two situations described in (H). \*\*\* $p < 0.001$  (Wilcoxon test). (J) Percentage of H3K9me3 domains overlapping ERVs in *Setdb1*<sup>-/-</sup> and *Suv39h1*<sup>-/-</sup> Th2 cells (relative to *Setdb1*<sup>+/+</sup> Th2 cells). All data are from two independent biological replicates for each genotype. See also Figure S5 and Table S3.



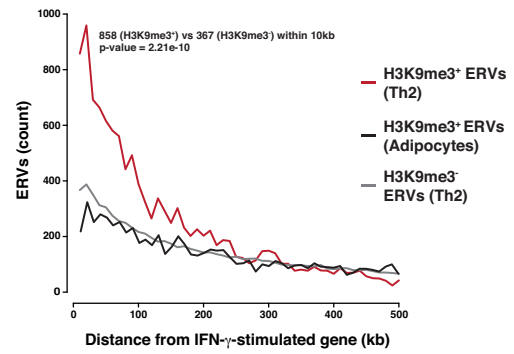
**Figure 5. SETDB1-deficiency upregulates a subset of ERVs and their neighboring genes.**

Analyses were focused on ERVs overlapping a SETDB1-dependent H3K9me3 domain in Th2 cells. ERVs overlapping gene bodies or promoters were excluded from the analysis. (A) Proportions and numbers of ERVs that were differentially ("Changed") or similarly ("Stable") expressed between *Setdb1*<sup>+/+</sup> and *Setdb1*<sup>-/-</sup> Th2 cells are shown on the left. The percentages of differentially expressed ERVs that were up-regulated ("Up") or down-regulated ("Down") are shown on the right. (B) Expression levels of all ERVs that are differentially expressed in *Setdb1*<sup>+/+</sup> and *Setdb1*<sup>-/-</sup> Th2 cells, as measured by RNA-seq. (C) Average ATAC-Seq signal profiles at ERVs that overlap ("H3K9me3<sup>+</sup>"), or not ("H3K9me3<sup>-</sup>"), with H3K9me3 domains in Th2 cells and that were transcriptionally upregulated ("H3K9me3<sup>+</sup> up", fold change >10, expression in *Setdb1*<sup>-/-</sup> ≥ 1), or not ("H3K9me3<sup>+</sup>"), in *Setdb1*<sup>-/-</sup> cells. The data represent the signal measured in *Setdb1*<sup>-/-</sup> Th2 cells relative to that measured in their wild-type counterparts. (D) Expression levels of the 238 enhancers located in the vicinity (± 5kb) of ERVs that overlap with H3K9me3 domains in *Setdb1*<sup>+/+</sup> cells and are more expressed in *Setdb1*<sup>-/-</sup> cells. The "0" on the X-axis corresponds to the center of the enhancer. \*\*\*p<0.001 Pearson's chi-squared test. (E) Average H3K4me1 or H3K27ac signal profiles at Th2 enhancers associated with H3K9me3<sup>+</sup> ERVs that are more expressed in *Setdb1*<sup>-/-</sup> Th2 cells. (F) GSEA of *Setdb1*<sup>-/-</sup> Th2 cells, with genes ranked on the basis of their expression in *Setdb1*<sup>-/-</sup> vs *Setdb1*<sup>+/+</sup> Th2 cells. The gene set comprises all the genes associated with enhancers located in the vicinity of ERVs that overlap with H3K9me3 domains and are more expressed in *Setdb1*<sup>-/-</sup> cells (left panel). A random selection of 500 genes associated with enhancers that remain silent in *Setdb1*<sup>-/-</sup> cells was used as control (right panel). (G) Gene expression change between *Setdb1*<sup>-/-</sup> and *Setdb1*<sup>+/+</sup> Th2 cells for genes that are expressed in at least one population and that are located in the vicinity of H3K9me3<sup>+</sup> ERVs that remain silent or become upregulated in SETDB1-deficient cells. \*p<0.05, \*\*\*p<0.001 (Wilcoxon test). All transcriptomic data are from three independent biological replicates. ChIP- and ATAC-seq data are from two independent biological replicates. See also Figure S6.

A

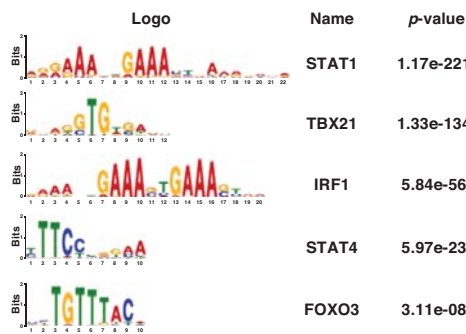


B



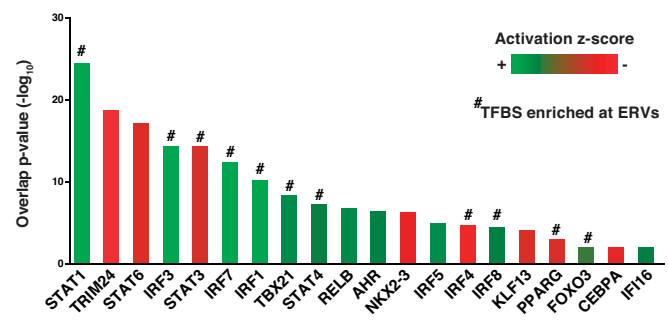
C

Pro-Th1 transcription factors with binding sites showing enrichment at H3K4me1-H3K9me3<sup>+</sup> ERV sequences

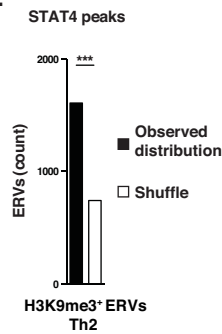


D

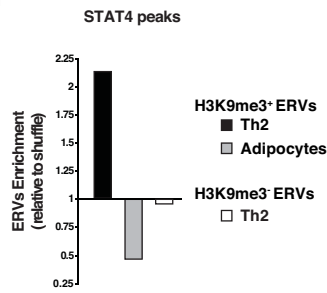
Transcriptional regulators that could explain changes in gene expression between *Setdb1*<sup>-/-</sup> and *Setdb1*<sup>+/+</sup> Th2 cells cultured in Th1 condition



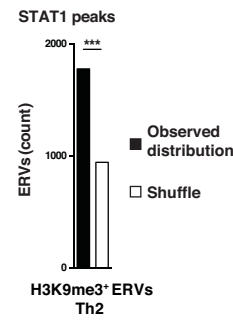
E



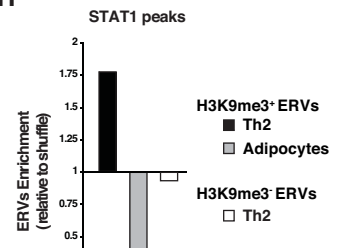
F



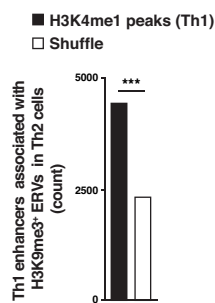
G



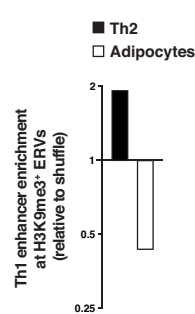
H



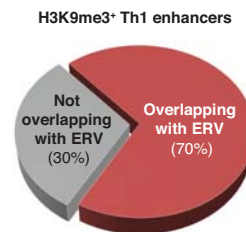
I



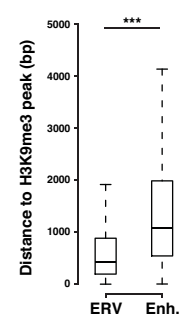
J



K

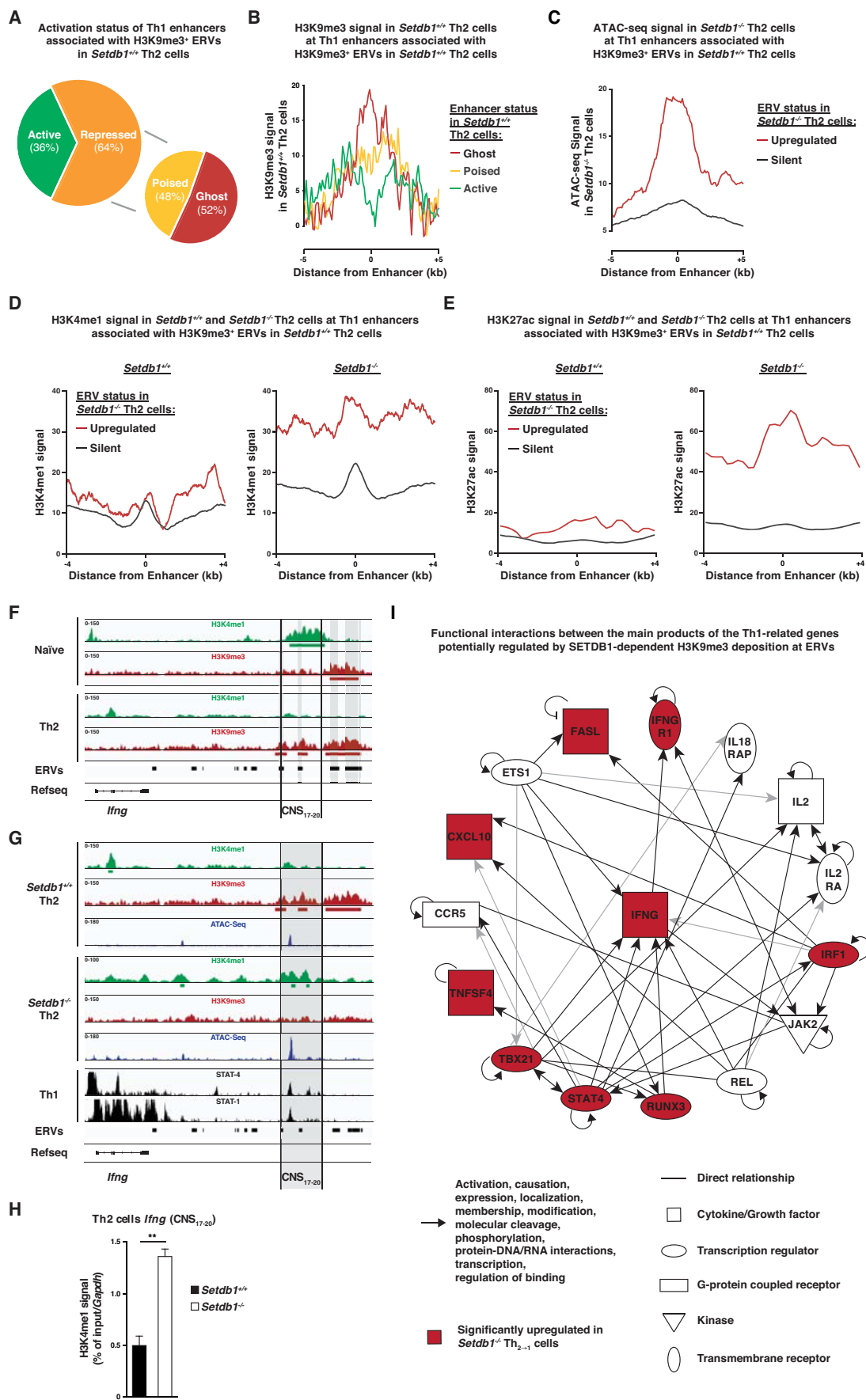


L



**Figure 6. H3K9me3<sup>+</sup> ERVs mark the Th1 enhancer landscape in Th2 cells.**

Only ERVs that do not overlap with Th2 enhancers (defined as H3K4me1<sup>+</sup> genomic regions) were analyzed. (A) Biological functions assigned to SETDB1-dependent H3K9me3<sup>+</sup> ERVs by GREAT. (B) Frequency histograms of absolute distances from each H3K9me3<sup>-</sup> (Th2) or H3K9me3<sup>+</sup> (Th2 or adipocytes) ERV to the nearest ISG. Statistical significance of the observed enrichment within the first 10 kb of the nearest ISG was assessed by chi-squared test. (C) Th1-related transcription factor binding sites showing enrichment within the 13,303 ERVs sequences, as determined using the AME module of the MEME suite. (D) The transcriptional regulators that might account for the differences in gene expression observed between *Setdb1*<sup>+/+</sup> and *Setdb1*<sup>-/-</sup> Th2 cells following culture in Th1-polarizing conditions were identified by using the Upstream Regulator Analysis module of the Ingenuity Pathway Analysis (IPA) software. The activation z-score infers the activation state of the transcriptional regulators in *Setdb1*<sup>-/-</sup> cells by comparing the differences in gene expression observed in *Setdb1*<sup>-/-</sup> and *Setdb1*<sup>+/+</sup> cells to the predicted effect (activating or inhibiting) of each transcriptional regulator on these genes. Hashtags highlight transcriptional regulators whose binding sites are enriched at ERVs. Only transcriptional regulators with an overlap p-value of less than 0.01 and a z-score lower than -2 or greater than 2 are shown. (E, G) Absolute numbers of H3K9me3<sup>+</sup> ERVs associated with Th1-specific STAT4 (E) or STAT1 (G) binding sites in Th2 cells (black bars) compared to counts obtained with a random distribution (shuffle, white bar). (F, H) Relative enrichment of H3K9me3<sup>+</sup> or H3K9me3<sup>-</sup> ERVs at Th1-specific STAT4 (F) or STAT1 (H) peaks in adipocytes and Th2 cells. (I) Numbers of putative Th1 enhancers, defined as Th1 H3K4me1<sup>+</sup> genomic regions with no overlap with promoters, associated with SETDB1-dependent H3K9me3<sup>+</sup>H3K4me1<sup>-</sup> ERVs in Th2 cells. (J) Relative enrichment of Th1 enhancers at H3K9me3<sup>+</sup> ERVs in adipocytes and Th2 cells. (K) Relative proportions of Th1 enhancers intersecting with H3K9me3 domains that overlap, or not, with ERVs in Th2 cells. (L) Box plots representing the distance between H3K9me3 domains and ERVs or Th1 enhancers. \*\*\**p*<0.001 (chi-squared test (B and E-J) or Wilcoxon test (L)). See also Figure S7.

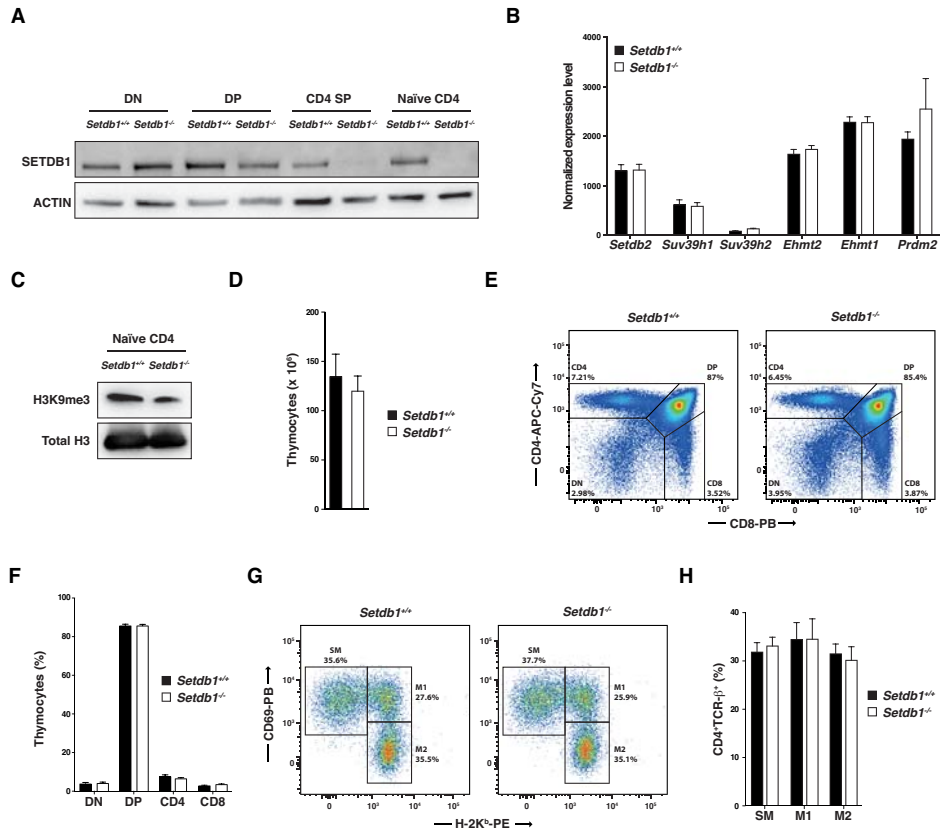




**Figure 7. SETDB1-dependent H3K9 trimethylation at ERVs represses Th1-specific enhancers.**

(A) Proportions of Th1 enhancers associated with H3K9me3<sup>+</sup> ERVs in Th2 cells that are either active (H3K4me1<sup>+</sup>H3K27ac<sup>+</sup>) or repressed (H3K4me1<sup>+</sup>H3K27ac<sup>-</sup> or H3K4me1<sup>-</sup>H3K27ac<sup>-</sup>). The percentages of repressed enhancers that are poised (H3K4me1<sup>+</sup>H3K27ac<sup>-</sup>) or “ghost” (H3K4me1<sup>-</sup>H3K27ac<sup>-</sup>) are shown on the right. (B) Average H3K9me3 signal profiles in *Setdb1*<sup>+/+</sup> Th2 cells at Th1 enhancers that are associated with H3K9me3<sup>+</sup> ERVs and that are either active, poised or repressed in *Setdb1*<sup>+/+</sup> Th2 cells. (C) Average ATAC-seq signal profiles in *Setdb1*<sup>-/-</sup> Th2 cells at Th1 enhancers that are repressed and associated with H3K9me3<sup>+</sup> ERVs in *Setdb1*<sup>+/+</sup> Th2 cells and that are more expressed, or not (“Silent”), in the absence of SETDB1. (D, E) Average H3K4me1 (D) and H3K27ac (E) signal profiles in *Setdb1*<sup>-/-</sup> and *Setdb1*<sup>+/+</sup> Th2 cells at Th1 enhancers that are repressed and associated with H3K9me3<sup>+</sup> ERVs in *Setdb1*<sup>+/+</sup> Th2 cells and that are upregulated, or not (“Silent”), in the absence of SETDB1. (F,G) Integrative Genomics Viewer (IGV) snapshots of the *Ifng* locus. The green, red and black tracks represent H3K4me1, H3K9me3 and STAT1/4 ChIP-seq signals, respectively. The blue tracks represent ATAC-seq signals. Colored boxes indicate peaks for H3K4me1 (green) and H3K9me3 (red). Black boxes correspond to ERV coordinates obtained by reconstruction as indicated in the Methods. (H) Enrichment of H3K4me1 at CNS<sub>17-20</sub> of the *Ifng* gene in *Setdb1*<sup>+/+</sup> and *Setdb1*<sup>-/-</sup> Th2 cells as measured by ChIP-qPCR. Data are means ± SEM of three independent biological replicates. (I) Th1 gene network generated with the IPA software shows the main Th1-related genes that have at least one enhancer in the vicinity (± 5kb) of a SETDB1-dependent H3K9me3<sup>+</sup>H3K4me1<sup>-</sup> ERV in Th2 cells.

**Figure S1**



**Figure S1. Normal intrathymic T cell development in *Setdb1*<sup>-/-</sup> mice. Related to Figure 1.**

(A) SETDB1 protein levels were determined by western blotting in CD4<sup>+</sup>CD8<sup>-</sup> (double-negative, DN), CD4<sup>+</sup>CD8<sup>+</sup> (double-positive, DP) and CD4<sup>+</sup>CD8<sup>-</sup> (CD4 single-positive, CD4 SP) thymocytes as well as in naïve CD4 T cells isolated from *Setdb1*<sup>+/+</sup> and *Setdb1*<sup>-/-</sup> mice. Actin was used as a control. Data are representative of three independent biological replicates.

(B) mRNA levels of various H3K9-specific methyltransferases in *Setdb1*<sup>+/+</sup> and *Setdb1*<sup>-/-</sup> naïve CD4 T cells as measured by RNA-seq. Data are represented as mean ± SD of four (*Setdb1*<sup>+/+</sup>) or three (*Setdb1*<sup>-/-</sup>) independent biological replicates.

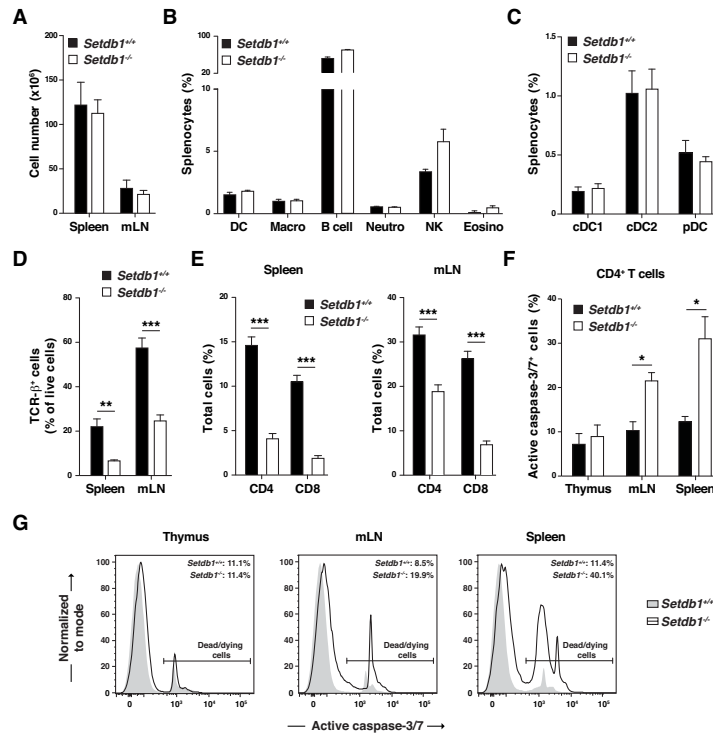
(C) H3K9me3 levels in *Setdb1*<sup>+/+</sup> and *Setdb1*<sup>-/-</sup> naïve CD4 T cells as determined by western blotting. Total H3 protein was used as a control. Data are representative of two independent biological replicates.

(D) Absolute numbers of live thymocytes in *Setdb1*<sup>+/+</sup> and *Setdb1*<sup>-/-</sup> mice. Data are represented as mean ± SEM of four independent experiments.

(E, F) CD4 and CD8 expression by *Setdb1*<sup>+/+</sup> and *Setdb1*<sup>-/-</sup> thymocytes as measured by flow cytometry. Representative dot-plots (E) as well as the relative proportions of the four main populations of thymocytes (F) are shown. The data are representative (E) or are represented as mean ± SEM (F) of four independent experiments.

(G, H) Relative proportions of the three main populations of CD4<sup>+</sup>TCR<sup>β</sup><sup>+</sup> thymocytes in *Setdb1*<sup>+/+</sup> and *Setdb1*<sup>-/-</sup> mice. The three populations were identified by flow cytometry based on H-2K<sup>b</sup> and CD69 expression. The CD69<sup>+</sup>H-2K<sup>b</sup> population is semi-mature (SM) and the CD69<sup>+</sup>H-2K<sup>b</sup> and CD69<sup>-</sup>H-2K<sup>b</sup> fractions correspond to the mature 1 (M1) and mature 2 (M2) populations, respectively (Xing, 2016). The data are representative (G) or are represented as mean ± SEM (H) of three independent experiments.

Figure S2



**Figure S2. Impaired peripheral T cell survival in *Setdb1*<sup>-/-</sup> mice. Related to Figure 1.**

(A) Absolute numbers of spleen and mesenteric lymph node cells in *Setdb1*<sup>+/+</sup> and *Setdb1*<sup>-/-</sup> mice.

(B) Relative proportions of dendritic cells (DC), monocytes/macrophages (Macro), B cells (B cell), neutrophils (Neutro), natural killer cells (NK) and eosinophils (Eosino) in *Setdb1*<sup>+/+</sup> and *Setdb1*<sup>-/-</sup> spleen, as measured by flow cytometry.

(C) Relative proportions of conventional type 1 (cDC1), conventional type 2 (cDC2) and plasmacytoid (pDC) dendritic cells in *Setdb1*<sup>+/+</sup> and *Setdb1*<sup>-/-</sup> spleens, as measured by flow cytometry.

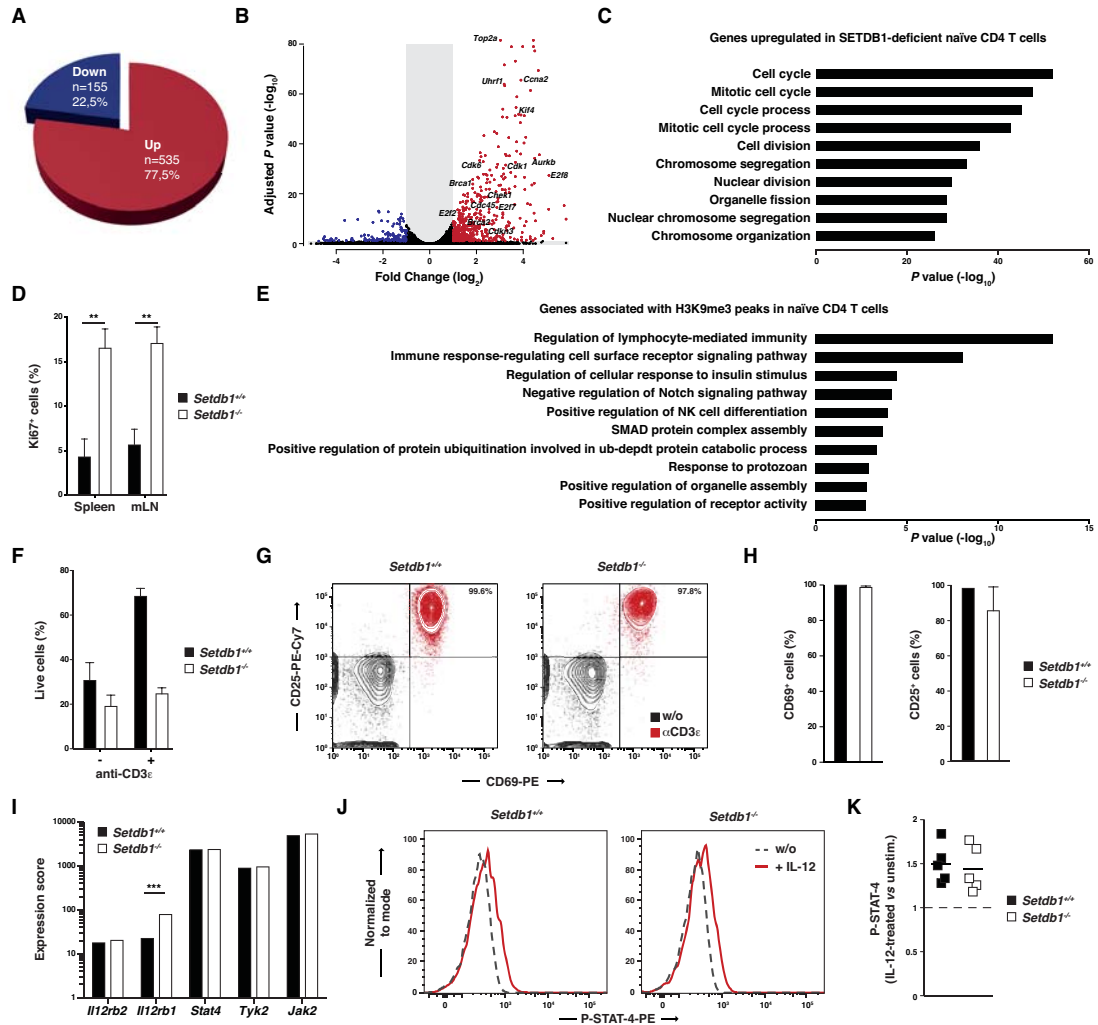
(D, E) Percentages of TCR- $\beta^+$  (D) and TCR- $\beta^+$ CD4<sup>+</sup> or TCR- $\beta^+$ CD8<sup>+</sup> (E) cells in the spleen and mesenteric lymph nodes from *Setdb1*<sup>+/+</sup> and *Setdb1*<sup>-/-</sup> mice, as measured by flow cytometry.

(F) Proportions of dead or dying CD4 T cells calculated from the data in (F). Data are represented as mean  $\pm$  SEM of three independent experiments.

(G) Percentages of dead or dying CD4 T cells in the thymus, spleen and mesenteric lymph nodes from *Setdb1*<sup>+/+</sup> and *Setdb1*<sup>-/-</sup> mice. Activated Caspases-3/7 were labelled with a fluorogenic substrate and staining was analyzed by flow cytometry. Data are representative of three independent experiments.

\* $p < 0.05$ , \*\* $p < 0.01$  and \*\*\* $p < 0.001$  (unpaired t-test).

**Figure S3**



**Figure S3. Normal activation of *Setdb1*<sup>-/-</sup> CD4 T cells despite impaired survival at early time points. Related to Figure 1.**

(A) Relative proportions and absolute number of genes that were upregulated (red) or downregulated (blue) in *Setdb1*<sup>-/-</sup> naïve CD4 T cells relative to *Setdb1*<sup>+/+</sup> naïve CD4 T cells, as measured by RNA-seq.

(B) Volcano-plot representing fold change (log<sub>2</sub>) in gene expression between *Setdb1*<sup>+/+</sup> and *Setdb1*<sup>-/-</sup> naïve CD4 T cells vs adjusted *P* value (-log<sub>10</sub>). Blue and red dots represent down- and up-regulated genes, respectively. Some of the genes implicated in the regulation of the cell cycle are highlighted.

(C) Gene ontology analysis of genes upregulated in *Setdb1*<sup>-/-</sup> naïve CD4 T cells when compared to *Setdb1*<sup>+/+</sup> naïve CD4 T cells. The data in parts (A–C) are from three (*Setdb1*<sup>-/-</sup>) or four (*Setdb1*<sup>+/+</sup>) independent biological replicates.

(D) Percentages of Ki67<sup>+</sup> naïve CD4 T cells in the spleen and mesenteric lymph nodes of *Setdb1*<sup>+/+</sup> and *Setdb1*<sup>-/-</sup> mice, as determined by flow cytometry.

(E) Gene ontology analysis of genes associated with H3K9me3 peaks in naïve CD4 T cells. The nearest gene to each peak was included in the data set. The data are from two independent biological replicates.

(F) Percentage of live *Setdb1*<sup>+/+</sup> and *Setdb1*<sup>-/-</sup> CD4 T cells following two days of culture in the absence (-) or presence (+) of an antibody against CD3ε.

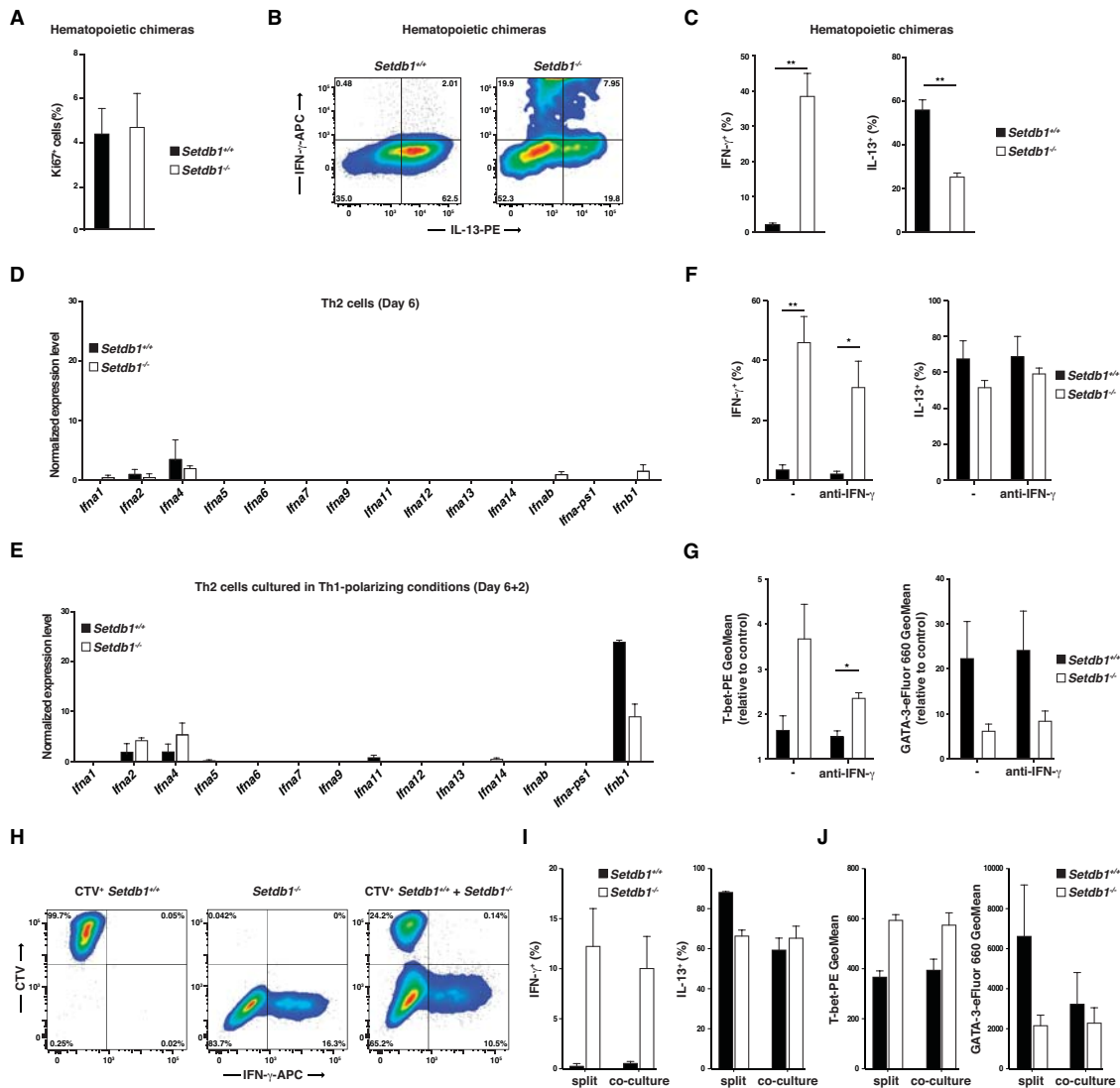
(G, H) Expression of the activation markers CD69 and CD25 by *Setdb1*<sup>+/+</sup> and *Setdb1*<sup>-/-</sup> CD4 T cells after two days of culture in the absence (black) or presence (red) of an antibody against CD3ε. Representative dot-plots of cell populations (G) and average percentage of CD69<sup>+</sup> and CD25<sup>+</sup> cells (H) are shown.

(I) mRNA levels of the main molecules involved in IL12-mediated signaling in *Setdb1*<sup>+/+</sup> and *Setdb1*<sup>-/-</sup> naïve CD4 T cells as measured by RNA-seq. Data are mean of four (*Setdb1*<sup>+/+</sup>) or three (*Setdb1*<sup>-/-</sup>) independent biological replicates.

(J, K) STAT4 phosphorylation level on tyrosine 693 in *Setdb1*<sup>+/+</sup> and *Setdb1*<sup>-/-</sup> CD4 T cells stimulated or not for 30 minutes with IL-12, as measured by flow cytometry. Representative histograms (J) and ratio of the geometric mean fluorescence intensity of stimulated sample and that of unstimulated control (K) are shown. Data are from 5 independent biological replicates from 2 independent experiment.

Data are represented as mean  $\pm$  SEM of three (D, F and measurements of CD69 in H) or four (measurements of CD25 in H) independent experiments. \*\**p*<0.01 and \*\*\**p*<0.001 (unpaired t-test).

**Figure S4**



**Figure S4. SETDB1 controls Th2 cell commitment through T cell-intrinsic mechanisms. Related to Figure 3.**

(A) Spleen cells were harvested from *Setdb1*<sup>+/+</sup> and *Setdb1*<sup>-/-</sup> mixed hematopoietic chimeras and the frequency of Ki67<sup>+</sup> cells in the CD45.2<sup>+</sup>CD90.2<sup>+</sup>CD4<sup>+</sup> T cell population was determined by flow cytometry. Data are mean  $\pm$  SEM of 3 independent experiments.

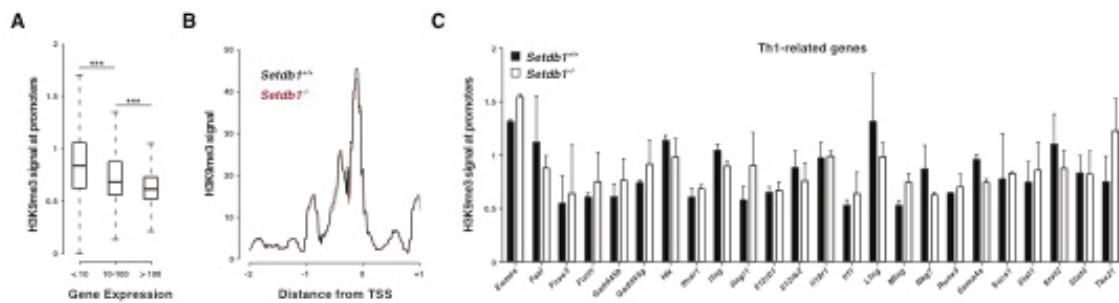
(B-C) CD45.2<sup>+</sup>CD90.2<sup>+</sup> naïve CD4 T cells were isolated from *Setdb1*<sup>+/+</sup> and *Setdb1*<sup>-/-</sup> mixed hematopoietic chimeras, cultured for 6 days in Th2 polarizing conditions, and then grown for 2 more days in Th1 medium. The percentages of IL-13- and IFN- $\gamma$ -producing cells were determined by flow cytometry. Data are mean  $\pm$  SEM (C) or are representative (B) of 3 independent experiments. \*\*p<0.01 (unpaired t-test).

(D, E) Levels of type I interferon mRNAs in *Setdb1*<sup>+/+</sup> and *Setdb1*<sup>-/-</sup> CD4 T cells after 6 days of culture in Th2-polarizing conditions (D), and after two further days of culture in Th1 medium (E). RNA-seq data are represented as mean  $\pm$  SEM of two (E, *Setdb1*<sup>+/+</sup>) or three independent experiments.

(F, G) *Setdb1*<sup>+/+</sup> and *Setdb1*<sup>-/-</sup> Th2 cells were cultured for four days in Th1 medium in the absence (-) or presence (anti-IFN- $\gamma$ ) of neutralizing antibody specific for IFN- $\gamma$ . The percentages of IL-13- and IFN- $\gamma$ -producing cells (F) as well as the expression of T-bet and GATA-3 (G) were determined by flow cytometry. Data are represented as mean  $\pm$  SEM of three independent experiments.

(H-J) *Setdb1*<sup>+/+</sup> Th2 cells were labeled with CellTrace Violet and cultured separately (split) or together with (co-culture) *Setdb1*<sup>-/-</sup> Th2 cells for 2 days in Th1 medium. The percentages of IFN- $\gamma$ - and IL-13-producing cells (H, I), as well as the expression of T-bet and GATA-3 (J), were determined by flow cytometry. Data are represented as mean  $\pm$  SEM or are representative of two independent experiments. \*p<0.05 and \*\*p<0.01 (unpaired t-test).

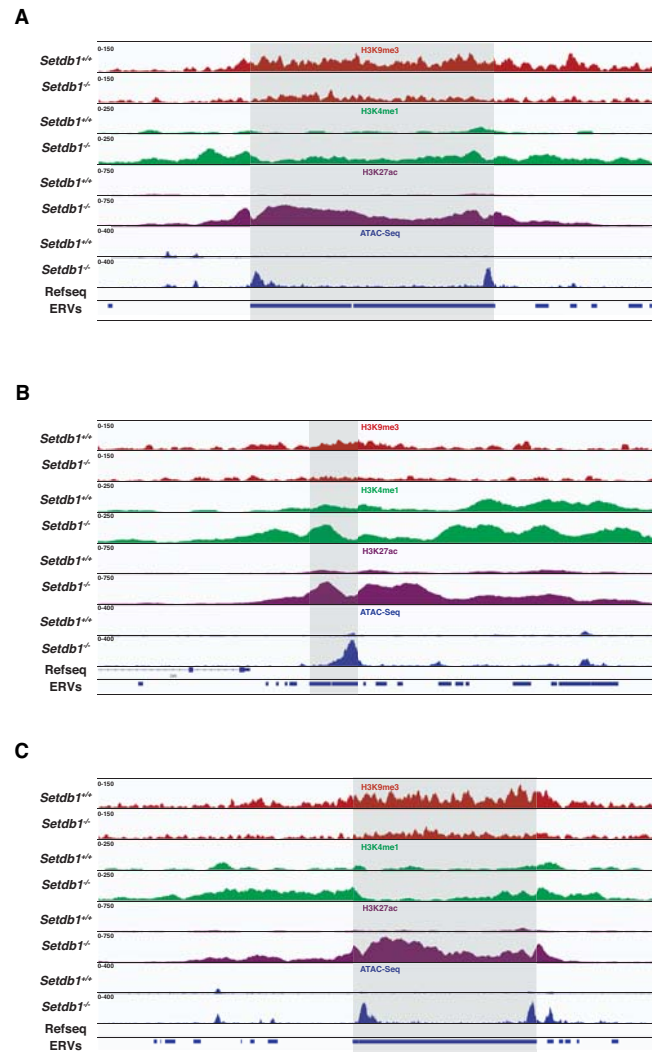
**Figure S5**



**Figure S5. SETDB1 does not seem to control CD4<sup>+</sup> T cell programming by modulating Th1 gene promoter activity. Related to Figure 4.**

(A) Box plot representing H3K9me3 signal at promoters of genes with expression levels below 10, comprised between 10 and 100, or above 100 in *Setdb1*<sup>+/+</sup> Th2 cells. The horizontal bar within each box corresponds to the median. \*\*\*p<0.001 (Wilcoxon test). (B) Average H3K9me3 signal at promoters of all genes in *Setdb1*<sup>+/+</sup> and *Setdb1*<sup>-/-</sup> Th2 cells. The "0" on the X-axis corresponds to the TSS. (C) H3K9me3 signal at promoters of Th1-related genes in *Setdb1*<sup>+/+</sup> and *Setdb1*<sup>-/-</sup> Th2 cells. Data are from two independent biological replicates.

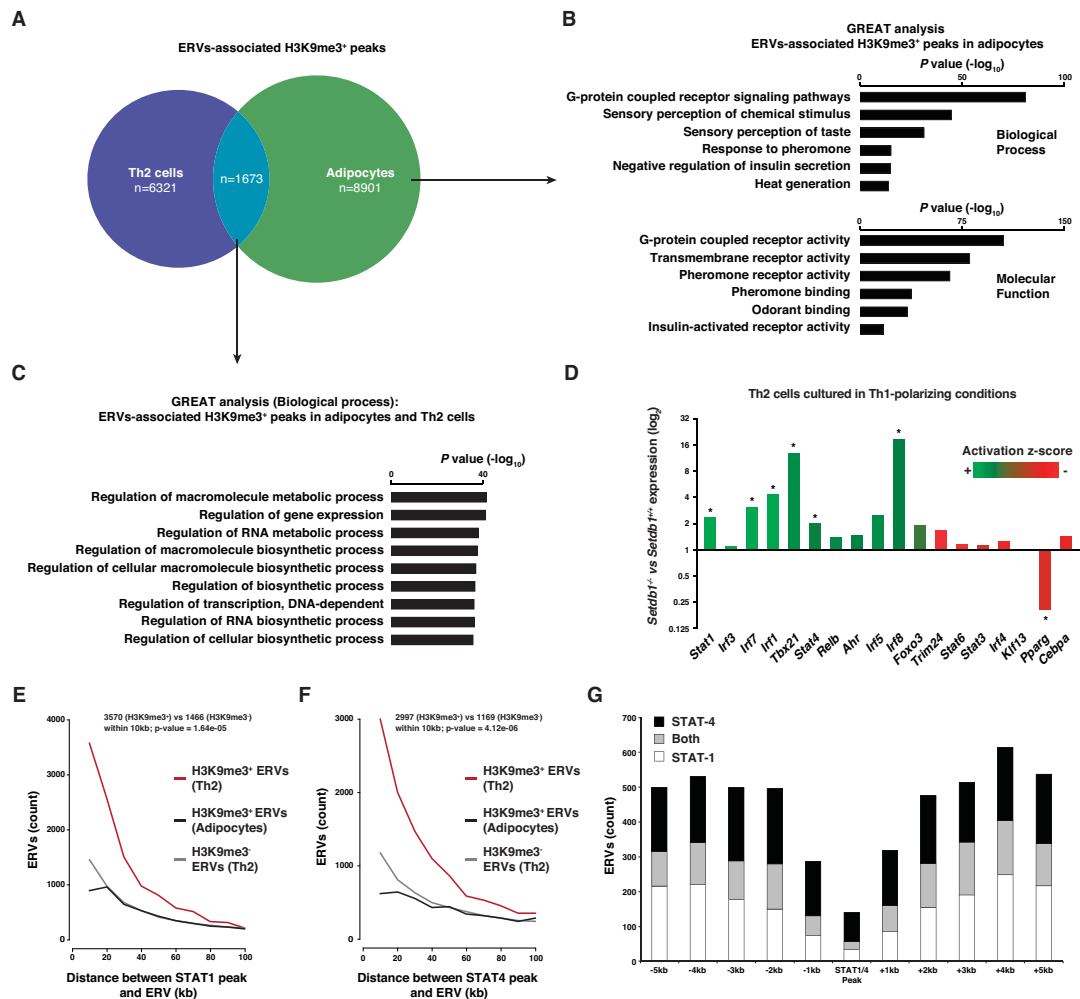
Figure S6



**Figure S6. Loss of H3K9me3 correlates with accumulation of permissive histone marks and increased chromatin accessibility at ERVs in *Setdb1*<sup>-/-</sup> Th2 cells. Related to Figure 5.**

(A-C) Integrative Genomics Viewer (IGV) snapshots of different genomic locations showing ERVs activation in *Setdb1*<sup>-/-</sup> Th2 cells. The red, green and purple tracks represent H3K9me3, H3K4me1 and H3K27ac ChIP-seq signals, respectively. The blue tracks represent ATAC-seq signals. Blue boxes correspond to ERV coordinates obtained by reconstruction as indicated in the Methods.

**Figure S7**



**Figure S7. The repertoire of H3K9me3<sup>+</sup> ERVs is cell-type specific. Related to Figure 6.**

(A) Venn diagram showing the number of H3K9me3 peaks overlapping ERVs in *Setdb1*<sup>+/+</sup> Th2 cells (blue) and/or in cells from white adipose tissue (green, Adipocytes). ChIP-seq data from adipocytes were obtained from the International Human Epigenome Consortium (IHEC, <http://ihc-epigenomes.org/>) and are merged from six independent biological replicates.

(B) Assignment of biological functions to H3K9me3 peaks overlapping ERVs in adipocytes but not in Th2 cells by analyzing the annotations of their nearby genes using GREAT.

(C) As in (B) but for H3K9me3 peaks overlapping ERVs in both adipocytes and Th2 cells.

(D) Ratio of expression of various genes encoding transcriptional regulators in *Setdb1*<sup>-/-</sup> compared to *Setdb1*<sup>+/+</sup> Th2 cells after culture in Th1-inducing conditions. The activation z-score was determined by performing an Upstream Regulator Analysis in IPA. It infers the activation state of the indicated transcriptional regulators in *Setdb1*<sup>-/-</sup> cells by comparing the differences in gene expression observed in *Setdb1*<sup>-/-</sup> versus *Setdb1*<sup>+/+</sup> cells to the predicted effect (activating versus inhibiting) of each transcriptional regulator on the differentially expressed genes.

(E-F) Frequency histograms of absolute distances from each H3K9me3<sup>+</sup> (Th2) or H3K9me3<sup>+</sup> (Th2 or adipocytes) ERV to STAT1 (E) or STAT4 (F) peaks. Statistical significance of the observed enrichment within the first 10 kb was assessed by chi-squared test.

(G) Absolute numbers of H3K9me3<sup>+</sup> ERVs in Th2 cells that overlap or flank Th1-cell specific STAT1 or STAT4 genomic locations (or both).



# MATERIEL ET METHODE

## Mice

Suv39h1-deficient mice were kindly provided by T. Jenuwein (206). The Setdb1 mutant mouse strain (common strain name EPD0028\_1\_B07; international strain designation B6Dnk;B6N-Setdb1tm1a(EUCOMM)Wtsi) was established as part of the International Mouse Phenotyping Consortium (EMMA ID: EM:04052) at the German Research Center for Environmental Health (Helmholtz Zentrum, Muenchen). The targeting vector was composed of the promoterless L1L2\_gt1 cassette inserted in the L3L4\_pZero\_kan plasmid backbone. The construct was microinjected into C57BL/6 ES cells (JM8.N4 parental cell line) and the L1L2\_gt1 cassette was inserted at position 95350414 of chromosome 3, upstream of Setdb1 exon 4. The cassette was composed of a lacZ–neomycin sequence flanked by Flp Recombinase Target (FRT) sites and followed by a loxP sequence. An additional loxP site was inserted downstream of Setdb1 exon 4 at position 95349598. Additional information on the Setdb1 mutant mouse strain can be found at <https://www.infrafrontier.eu/search?keyword=EM:04052>. Mice with a conditional ready Setdb1 allele (Setdb1fl) were generated by intercrossing Setdb1 mutant mice with mice expressing the Flipper recombinase under the control of the ubiquitous Rosa26 promoter. Conditional knockout mice (Setdb1<sup>-/-</sup>) were obtained by intercrossing Setdb1fl/fl and CD4-CRE mutant mice. All the mice were bred and housed at the Regional Centre of Functional Exploration and Experimental Resources (CREFRE, UMS006/INSERM). Sex-matched 6- to 12-week-old mice were used and compared in all experiments. All experiments involving animals were conducted according to animal study protocols approved by the local ethics committee (# 16-U1043-JVM-496 and 16-U1043-JVM-20).

## Naïve CD4<sup>+</sup> T cell isolation

Spleen and lymph nodes (mesenteric, inguinal, axillary, brachial and cervical) were collected and digested with Liberase TM and DNase I (Sigma). Single-cell suspensions were then pooled and depleted of erythrocytes by osmotic shock (Red Blood Cell Lysis buffer, Sigma). CD4 T cells were enriched by negative selection by using antibodies specific for CD16/32 (2.4G2), I-A/I-E (M5/114.15.2), CD8 $\alpha$  (H59) and B220 (RA3-6B2), and Dynabeads sheep anti-rat IgG (Thermo Fisher Scientific). Naïve CD4 T cells, defined as CD4<sup>+</sup>CD25<sup>-</sup>CD62L<sup>high</sup>CD44<sup>low</sup>, were labeled with fluorochrome-conjugated monoclonal antibodies specific for CD4 (GK1.5, BD Biosciences), CD25 (PC61, BD Biosciences), CD62L (MEL14, Thermo Fisher Scientific) and

CD44 (IM7, BD Biosciences), and purified from the enriched fraction of CD4 T cells by fluorescence-activated cell sorting (FACS Aria, BD Biosciences).

### **T helper cell cultures**

Naïve CD4 T cells were cultured for three days in 96-well flat bottom plates coated with 10 µg/mL anti-CD3 $\epsilon$  antibody (145-2C11, InVivoMab™, BioXcell) in RPMI 1640 Glutamax™ supplemented with 1 mM sodium pyruvate, non-essential amino acids, 10 mM HEPES, 100 units/mL penicillin, 100 µg/mL streptomycin, 50 mM 2 $\beta$ -mercaptoethanol, 10% fetal calf serum (all from Thermo Fisher Scientific) and 1 µg/mL anti-CD28 antibody (37.51, InVivoMab™, BioXcell). Unless stated otherwise, Th1 medium also contained 10 ng/mL recombinant mouse IL-12 (R&D Systems) and 10 µg/mL anti-IL-4 neutralizing antibody (11B11, InVivoMab™, BioXcell). Th2 medium contained 50 ng/mL recombinant mouse IL-4 (R&D systems) and 10 µg/mL anti-IFN- $\gamma$  neutralizing antibody (XMG1.2, InVivoMab™, BioXcell). At day 3, the cells were re-plated in the same conditioning medium but without the anti-CD3 $\epsilon$  and anti-CD28 antibodies and with 30 IU/mL recombinant IL-2 (Proleukin). To test for Th2 cell lineage commitment, cells were harvested at day 6, extensively washed in complete medium, and re-plated in Th1-polarizing conditions as indicated above. To assess the role of the IFN- $\gamma$  pathway in Th2 cell plasticity, Th1 medium was supplemented with 10 µg/mL anti-IFN- $\gamma$ . In co-culture experiments, Setdb1<sup>+/+</sup> and Setdb1<sup>-/-</sup> Th2 cells were differentiated separately, mixed at a 1:3 ratio, and then plated in Th1 culture conditions.

### **T cell proliferation and differentiation analysis by flow cytometry**

To analyze intracellular transcription factor expression upon T helper cell differentiation, cells were collected at the requires time points, stained with the fixable viability dye eFluor 506 (Thermo Fisher Scientific), and labeled with fluorochrome-conjugated antibodies specific for T-bet (ebio4B10, Thermo Fisher Scientific) and GATA-3 (TWAJ, Thermo Fisher Scientific) by means of the Transcription Factor Staining Buffer Set (Thermo Fisher Scientific). For intracellular cytokine staining, cells were first stimulated at 37°C with 20 ng/mL phorbol 12-myristate 13-acetate (Millipore) and 1 µg/mL ionomycin (Millipore) for 5 hours in the presence of GolgiStop™ (BD Biosciences). Cells were then labelled with the fixable viability dye eFluor 506 and stained with fluorochrome-coupled antibodies specific for IL-13 (ebio13A, Thermo Fisher Scientific), IFN- $\gamma$  (XMG1.2, Thermo Fisher Scientific), GM-CSF (MP1-22E9, BD Biosciences), TNF (MP6-XT22, Thermo Fisher Scientific) or IL-4 (11B11, BD Biosciences) by using the Cytofix/Cytoperm™ Fixation/Permeabilization Kit (BD Biosciences). When indicated, naïve CD4 T cells were labeled prior to culture with 0.5 µM CellTrace Violet (Thermo Fisher Scientific). Flow cytometry was performed by using a LSRII Fortessa cytometer (BD Biosciences) or

MACSQuant analyzer 10 (Myltenyi) and the data were analyzed by using FlowJo software (Tree Star).

### **Mouse phenotyping**

To determine the frequency and phenotype of immune cells in primary and secondary lymphoid organs, thymus, spleen and lymph nodes were collected from Setdb1<sup>-/-</sup> and Setdb1<sup>+/+</sup> mice and single-cell suspensions were obtained by mechanical disruption. Cells were then incubated on ice in FACS buffer (PBS, 3 mM EDTA, 3% fetal calf serum) containing 10 µg/mL anti-CD16/32 antibody for 20 minutes. Fluorochrome-conjugated antibodies were added at saturating concentrations, and cell suspensions were incubated on ice and protected from light for a further 20 minutes. For intracellular staining, cells were fixed and permeabilized by using the Transcription Factor Staining Buffer Set (Thermo Fisher Scientific) according to the manufacturer's instructions. The following antibodies were used for phenotyping: anti-TCR-β (H57-597), anti-CD4 (GK1.5), anti-NKp46 (29A1.4), anti-CD11b (M1/70), anti-CD19 (1D3), anti-CD25 (PC61), anti-CD69 (H1.2F3), anti-Siglec-F (E50-2440), anti-H-2Kb (AF6-88.5) and anti-Ki67 (B56) all from BD Biosciences; anti-CD8β (ebioH35-17.2), anti-CD8α (53-6.7), anti-PDCA1 (ebio927), anti-I-A/I-E (M5/114), anti-CD11c (N418), anti-Gr1 (RB6-8C5), anti-B220 (RA3-6B2), anti-CD44 (IM7) and anti-CD62L (MEL-14), all from Thermo Fisher Scientific. Dendritic cells (DC) were gated based on I-Ab and CD11c expression (CD19-TCR-β-CD11c+I-Ab+) and CD8α, CD11b and PDCA-1 were used as markers to identify the conventional type 1 (cDC1, CD8α+CD11b-), conventional type 2 (cDC2, CD8α-CD11b+) and plasmacytoid (pDC, PDCA-1+) sub-populations. Monocytes/macrophages (Macro) and B cells (B cell) were defined as lin-CD11c-CD11b+SSC-A<sub>low</sub>Gr-1<sub>low</sub>- and TCR-β-CD19+B220+, respectively. Neutrophils (Neutro), natural killer cells (NK) and eosinophils (Eosino) were identified as lin-CD11c-CD11b+SSC-A<sub>high</sub>Gr-1+, TCR-β-NKp46+ and lin-CD11c-CD11b+SSC-A<sub>high</sub>Gr-1-, respectively. Flow cytometry was performed by using a LSRII Fortessa cytometer (BD Biosciences) and the data were analyzed by using FlowJo software (Tree Star).

### **Ex vivo measurement of apoptosis**

Single-cell suspensions of spleen, mesenteric lymph nodes and thymus were obtained as described above. Spleen and lymph node cells were labeled with antibodies specific for TCR-β (H57-597, Thermo Fisher Scientific) and CD4 (GK1.5, BD Biosciences), whereas thymocytes were stained with antibodies specific for CD4 and CD8β (ebioH35-17.2, ThermoFisher Scientific). Apoptotic cells were then labeled by using the Cell Event™ Caspase-3/7 Green Detection Reagent (Thermo Fisher Scientific) according to the manufacturer's instructions. Flow cytometry was performed by using a LSRII Fortessa cytometer (BD Biosciences) and the data were analyzed by using FlowJo software (Tree Star).

### **Measurement of cytokines in cell culture supernatants**

Following 6 days of culture in Th1 or Th2 polarizing conditions, T cells were collected and extensively washed in complete medium. The differentiated cells ( $7.5 \times 10^4$  per well) were cultured overnight in 96-well flat bottom plates coated with anti-CD3 $\epsilon$  antibody in complete culture medium containing anti-CD28 antibody. The concentrations of cytokines in the cell culture supernatants were then measured by flow cytometry using the FlowCytomix Kit (a bead-based multiple cytokines detection system) according to the manufacturer's instructions (FlowCytomix, eBiosciences). Flow cytometry was performed by using a MACSquant Q10 flow cytometer (Miltenyi).

### **Phospho-STAT4 (Tyr693) staining**

Naïve CD4 T cells were cultured for three days in 96-well flat bottom plates in complete medium supplemented with 30 UI/mL of recombinant IL-2 (Peprotech) at a ratio of 1 to 1 with Dynabeads Mouse T-Activator CD3/CD28 (Thermo Fisher Scientific). After 3 washes in complete medium,  $10^5$  cells were stimulated for 30 minutes with recombinant mouse IL-12 (10 ng/mL). Cells were then fixed with BD Phosflow™ Lyse/Fix Buffer, permeabilized with BD Phosflow™ Perm Buffer III in the presence of 10mM Sodium Fluoride and 1mM Sodium Orthovanadate (Sigma), and stained with fluorochrome-conjugated monoclonal antibodies specific for phosphorylated STAT4 (pY693, clone 38/pSTAT4, BD). Flow cytometry was performed by using a LSRII Fortessa cytometer and the data were analyzed by using FlowJo software.

### **Mixed hematopoietic bone marrow chimeras**

Bone marrow from femurs and tibias was collected from CD45.2+CD90.1+ C57BL/6 mice and from Setdb1+/+ or Setdb1-/- CD45.2+CD90.2+ littermates. Single-cell suspensions were washed in complete medium. CD90.1+ and CD90.2+ cells were then eliminated using HO-22 and AT83 hybridoma supernatants, respectively, and rabbit complement (TCS Biosciences). Following treatment with DNase I (Sigma), single-cell suspensions were washed three times in PBS and filtered through a 70  $\mu$ m cell strainer. CD45.2+CD90.1+ cells were mixed at a ratio of 3 to 7 with either Setdb1+/+ or Setdb1-/- CD45.2+CD90.2+ cells.  $10^7$  cells were then injected intravenously into  $\gamma$ -irradiated CD45.1+ C57BL/6 hosts (11 Gy;  $^{137}\text{Cs}$  source) that were kept on antibiotic-containing water (0.28% pediatric suspension of Bactrim; Roche) for the next 4 weeks. Spleen and lymph nodes were isolated from these mice at least 6 weeks after injection of the bone marrow cells.

## Immunization

Mice were immunized subcutaneously on each side of the base of the tail with 20 µg of 1W1K peptide (EAWGALANKAVDKA, Genecust) in 100 µL of RIBI adjuvant (Sigma). 8 days after immunization, the lymph nodes draining the site of immunization were collected and homogenized. To determine the frequency of Ag-specific CD4<sup>+</sup> T cells, 4.10<sup>6</sup> cells were incubated with the 1W1K-I-Ab tetramer (NIH tetramers core facility) for 2h at room temperature. Cells were then labelled with fluorochrome-conjugated monoclonal antibodies specific for CD4 (RM4.5), CD44 (IM7), CD8α (53-6.7) and B220 (RA3-6B2) and with the fixable viability dye eFluor 506 (Thermo Fisher Scientific) according to the manufacturer's instructions. To determine the frequency of antigen-specific CD4<sup>+</sup> T cells producing IFN-γ, 4.10<sup>6</sup> lymph node cells were stimulated overnight with 7µM of the 1W1K peptide in complete medium and in the presence of Brefeldin A (5µg/mL, Sigma) and GolgiStop™ (BD Biosciences) during the last 5h. Cells were then harvested and stained with fluorochrome-conjugated monoclonal antibodies specific for CD4 and CD44 and with the fixable viability dye eFluor 506. Following fixation and permeabilization (Cytofix/Cytoperm™ Fixation/Permeabilization Kit, Thermo Fisher Scientific), cells were finally stained with fluorochrome-coupled antibodies specific for IFN-γ. To determine the Th1/Th2 balance of the antigen-specific T cell response, 2.10<sup>6</sup> cells were stimulated for 72h with 7µM of 1W1K peptide in HL-1 medium (Lonza) supplemented with 2 mM of glutamine, 100 units/mL penicillin and 100 µg/mL streptomycin. IL-13 and IFN-γ levels were then measured by ELISA in the cell culture supernatants.

## Western blotting

The different subpopulations of thymocytes were sorted on a FACS Aria (BD Biosciences) based on their expression of CD4 and CD8. Naïve CD4 T cells were purified from the spleen and lymph nodes as described above. Cells were lysed in 1X NuPAGE LDS sample buffer and 1X NuPAGE sample reducing agent (Thermo Fisher Scientific). Whole cell lysates were then sonicated briefly and proteins were separated by SDS-PAGE on 4–12% Bis-Tris gels (Thermo Fisher Scientific), transferred onto nitrocellulose membranes (BA-S 83 Optitran, GE Healthcare Life Sciences), and probed with antibodies specific for SETDB1 (ab107225, Abcam), total H3 (ab1791, Abcam), H3K9me3 (D4W1U, Cell Signaling Technology) or beta ACTIN (ab8227, Abcam). The bands were detected by using Amersham ECL western blotting detection reagent (GE Healthcare Life Sciences) and the ChemiDoc XRS+ imaging system (Bio-Rad) after staining with secondary antibodies coupled to horseradish peroxidase. Images were analyzed by using Image Lab software (Bio-Rad).

### **RNA-seq sample preparation and analysis**

Total RNA was extracted by using the RNeasy Micro Kit (Qiagen) and its quality was assessed on a 2100 Bioanalyzer (Agilent Technologies). Only high-quality RNA (i.e. RNA of integrity number > 7) was subsequently used to prepare the libraries according to the ScriptSeq RNA-seq protocol (Illumina). Quality controls of the libraries were performed by using standard methods, including quantification by Qubit (Thermo Fisher Scientific) and assessment of size distribution by using the 2100 Bioanalyzer. Samples were indexed and sequenced on an Illumina HiSeq 2500 or 3000 (paired-end reads of 100 or 150 bp, respectively). After trimming of adaptor sequences (Cutadapt 1.3) and removal of low-quality bases ( $-q$  value, < 15), high-quality reads were aligned to the mouse reference genome mm10 (Genome Reference Consortium) by using TopHat version 2.0.5 (296). Count of the reads mapping to each gene was performed using Htseq-count. Differential expression analysis was performed by using the DESeq package (Bioconductor software) (Anders and Huber, 2010). An adjusted P value of < 0.1 (P value adjusted for multiple testing with the Benjamini–Hochberg procedure) was used as cutoff to select the genes differentially expressed.

### **ChIP, semi-quantitative PCR and library preparation and sequencing**

ChIP was performed as previously described (297). Briefly, following cell lysis, the chromatin was sonicated with a Bioruptor Pico (Diagenode) to obtain fragments of 100-300 bp. In each assay, we used 5-50 million cells and 2-10  $\mu$ g of antibody specific for H3K9me3 (ab8898, Abcam), H3K27ac (Ab4729, Abcam) or H3K4me1 (ab8895, Abcam). Immunoprecipitation was performed by using Dynabead® Protein G (Thermo Fisher Scientific). Library preparation was carried out by using the TruSeq ChIP Sample Preparation Kit (Illumina). Library quality was assessed by using the 2100 Bioanalyzer and sequencing was performed on an Illumina HiSeq 3000 (paired-end reads of 150 bp). When indicated, semi-quantitative PCR was performed on a Light Cycler® 480 (Roche) using LightCycler 480 SYBR Green I Master (Roche). Primers specific for Ifng CNS17-20 (forward: tcctagactctgccactct; and reverse: gctcacatcaataggcgtg) and for the glyceraldehyde-3-phosphate dehydrogenase (Gapdh) promoter (forward: gctccttgcccttcagatt and reverse: cccttcccaccctgttcac) were used. The results were expressed as the percentage of input DNA normalized to the signal from the Gapdh promoter. More details on STAT1, STAT4 and H3K4me1 ChIP-seq data in Th1 cells can be found in (62).

### **ChIP-seq data processing**

Reads were filtered as described for RNA-seq and aligned to the mm10 reference genome by using BWA v.0.7.10 (298). H3K4me1 and H3K27ac peaks were detected by using the ‘broad’ option of MACS2 v.2.1.0 (299). To detect H3K9me3 domains, we used the R Bioconductor package CSAW v.1.4.1 (300). The minimum mapping quality score and the FDR threshold were

set to 50 and 0.05, respectively. We tested 11 window sizes ranging from 200 to 600 bp. In the end, we selected 300 bp as it allowed for the most accurate detection of H3K9me3-enriched domains, as determined using the visualization tool IGV. Differential binding windows were clustered in regions with the 'mergeWindows' function and the Benjamini-Hochberg method was applied to control the False Discovery Rate (FDR) across all detected clusters ('combineTests' function). For visualizing H3K4me1 and H3K27ac signals in Setdb1+/+ and Setdb1-/- Th2 cells, reads were extended by 200 bp because of low reads coverage.

### **ATAC-seq sample preparation and analysis**

ATAC-seq was performed as previously described (301) with some modifications. Briefly, 50 000 cells were lysed in ice-cold lysis buffer and the transposition reaction was performed using the Tn5 transposase at 37 °C for 30 min. DNA was purified using the Qiagen MinElute kit (Qiagen). The libraries were prepared using the Nextera DNA Library Prep Kit (Illumina) and purified using AMPure XP beads (Beckman) following a double-sided protocol to remove primer dimers and large fragments. Samples were performed in duplicates, multiplexed and sequenced on NextSeq-500 (75 bp paired-end reads) at the Transcriptomic & Genomic Platform Marseille Luminy (TGML, Marseille, France). Reads were aligned with BWA mem (version 0.7.12-r1039). Samples were normalized by scaling reads numbers according to the signal found at promoter of genes which were highly (expression level > 5000) and equally (0.9 < adjusted p-value) expressed in Setdb1+/+ and Setdb1-/- samples, as determined by RNA-seq.

### **ERV reconstruction**

Annotations of ERV elements were downloaded from the UCSC Genome Browser (assembly GRCm38, release of RepeatMasker: 2012-02-07). We used the four major subfamilies (ERV1, ERVK, ERVL and ERVL-MaLR) of LTRs and excluded elements for which the curator was unsure of the classification. We merged ERV fragments from the same family (identical 'repName') into a single ERV when located within 20 bp, as previously described (302). Count of the reads mapping to each ERV was performed using Htseq-count (303) and normalization was performed with DESeq. ERVs with an expression score  $\geq 1$  were considered as expressed.

### **Bioinformatics analyses**

R (<https://www.R-project.org>), SAMtools (Li et al., 2009) and the BEDtools suite v2.22.1 (304) were used to analyze high-throughput sequencing files. To determine the genome-wide distribution of H3K9me3 domains, we defined the different genomic regions as follows: gene body coordinates were extracted from assembly GRCm38; promoters were defined as transcription start sites +1kb/-2kb; enhancers were identified as H3K4me1 peaks with no



overlap with promoters; ERV coordinates were rebuilt from the RepeatMasker database as described above. As a control, we randomly distributed H3K9me3 domains through the genome using the shuffle sub-command of the BEDtools suite. R package 'Genomation' was used to visualize genomic intervals (305). Biological functions analysis of H3K9me3 ChIP-seq peak coordinates was performed by using the Genomic Regions Enrichment of Annotations Tool (GREAT) (306) with default settings and using the 'single nearest gene' option (each gene is assigned a regulatory domain that extends in both directions (within 100kb) to the midpoint between the gene's TSS and the nearest gene's TSS but no more than the maximum extension in one direction). For Gene Ontology analysis, the "enrichment analysis" tool from the Gene Ontology Consortium was used (<http://geneontology.org>). For analysis of motif enrichment, we used AME software of the MEME Suite version 4.11.3 with default options (0.05  $\geq$  adjusted p-value) (307); the HOCOMOCOv11 MOUSE was used as the input transcription factor motif database. We also used the gene set enrichment analysis (GSEA) software (<http://software.broadinstitute.org/gsea/index.jsp>) with default settings except for the 'Collapse dataset to gene symbols' and 'the permutation type' which were set as 'false' and 'gene set', respectively. Heat maps were generated by using matrix2png version 1.2.1 (<http://www.chibi.ubc.ca/matrix2png>). To measure H3K9me3, H3K4me1 or H3K27ac signal, we used bamCoverage from the deepTools suite (v2.3.4) to generate normalized bedgraph files (we normalized by sequencing depth and ignored chrX, chrY and chrM) and Bedtools map to calculate the average signal. Correlation analysis between H3K9me3+ or H3K9me3- ERVs and IFN- $\gamma$ -stimulated genes (ISG) was performed as described elsewhere (294). The absolute distance between ERVs and the nearest ISG (n=4,276) was first determined for all 13,303 H3K9me3+H3K4me1- retroelements. The distances were then grouped by 10 kb bin sizes. As control, a similar analysis was performed using an equal number of randomly sampled H3K9me3- ERVs. Sampling was repeated 10 times and the mean number of elements was used. Statistical significance was determined for the first 10 kb bin by chi-squared test.

## QUANTIFICATION AND STATISTICAL ANALYSIS

Statistical parameters including the exact value and significance of n and precision measures (Mean +/- SEM or SD) as well as statistical significance are reported in the figures and figure legends. Unless stated otherwise, asterisks denote statistical significance as calculated by Student's t test in GraphPad PRISM 6 (\*, p < 0.05; \*\*, p < 0.01; \*\*\*, p < 0.001). When large sets of unpaired data were compared, Pearson's chi-squared test was calculated in R to determine whether the observed difference between the sets of data arose by chance.



## DATA AVAILABILITY

### Data Resources

Raw and processed data files from ChIP-seq, ATAC-seq and RNA-seq experiments have been deposited in the NCBI Gene Expression Omnibus (<http://www.ncbi.nlm.nih.gov/geo/>) under accession number GSE101546.

## SUPPLEMENTAL TABLES

**Table S1.** Related to Figure 1. mRNA levels of genes related to T helper cell activation and differentiation in Setdb1<sup>+/+</sup> and Setdb1<sup>-/-</sup> naïve CD4 T cells, as measured by RNA-seq.

| gene_name | baseMean WT | baseMean KO | foldChange  | log2FoldChange | padj        |
|-----------|-------------|-------------|-------------|----------------|-------------|
| Ets1      | 50362.85972 | 41115.01425 | 0.816375688 | -0.292694876   | 0.427862878 |
| Stat1     | 18509.91519 | 29493.50535 | 1.593389545 | 0.672099014    | 0.134336564 |
| Jak1      | 18139.97019 | 20663.81563 | 1.13913173  | 0.187934592    | 1           |
| Itk       | 17511.86327 | 14735.40441 | 0.841452687 | -0.249045941   | 0.715686846 |
| Hsp90b1   | 13277.98599 | 12583.5966  | 0.947703711 | -0.077492007   | 1           |
| Pik3cd    | 11036.22517 | 14991.54916 | 1.358394644 | 0.441902675    | 0.083586378 |
| Lef1      | 10311.55171 | 9960.945913 | 0.965998736 | -0.049906794   | 1           |
| Tgfb2     | 9128.563878 | 9771.353466 | 1.070415193 | 0.098170498    | 1           |
| Prkcq     | 8470.696305 | 8589.038334 | 1.013970756 | 0.020016044    | 1           |
| Ptpn6     | 8433.165584 | 7623.001355 | 0.903931184 | -0.145715149   | 1           |
| Cd28      | 7828.500833 | 5774.966872 | 0.737684902 | -0.438923386   | 0.054676134 |
| Vav1      | 6716.938778 | 6285.495421 | 0.935767859 | -0.095777417   | 1           |
| Il4ra     | 6310.871592 | 7649.571155 | 1.212125939 | 0.277539602    | 0.639123561 |
| Itgb2     | 6302.010304 | 7917.947865 | 1.256416204 | 0.329314455    | 0.394971122 |
| Irf1      | 6184.958432 | 7604.409566 | 1.229500513 | 0.298072337    | 0.431753614 |
| Mbd2      | 6083.856225 | 6059.190662 | 0.995945735 | -0.005860956   | 1           |
| Junb      | 5660.654252 | 5698.537968 | 1.006692462 | 0.009623017    | 1           |
| Cdkn1b    | 5410.360308 | 4380.970388 | 0.80973727  | -0.304474212   | 0.450362216 |
| Jak2      | 4969.258967 | 5379.517443 | 1.082559287 | 0.114446037    | 1           |
| Nfkb1     | 4575.23706  | 5927.981581 | 1.295666542 | 0.373694468    | 0.204489952 |
| Cd69      | 4211.478088 | 6274.513965 | 1.48986029  | 0.57517705     | 0.513278502 |
| Stat6     | 4094.913866 | 4174.738856 | 1.019493692 | 0.027852849    | 1           |
| Stat3     | 4070.225599 | 5269.362625 | 1.294611932 | 0.372519706    | 0.150445566 |
| Mapk1     | 4040.340149 | 4743.352512 | 1.173998311 | 0.231430333    | 0.809534904 |
| Nfatc3    | 3913.667626 | 6044.376975 | 1.544427773 | 0.627072403    | 0.325978308 |
| Il17ra    | 3865.671491 | 4201.1999   | 1.086796928 | 0.120082393    | 1           |
| Lamp1     | 3807.746258 | 3855.948747 | 1.01265906  | 0.018148533    | 1           |
| Rictor    | 3510.495161 | 3938.326571 | 1.121872098 | 0.165908207    | 1           |
| Stat5b    | 3510.405662 | 3482.768943 | 0.992127201 | -0.011402994   | 1           |

|         |             |             |             |              |             |
|---------|-------------|-------------|-------------|--------------|-------------|
| Itgal   | 3373.941056 | 4954.828458 | 1.468558097 | 0.554400341  | 0.028154358 |
| Nbr1    | 3213.451559 | 3262.48338  | 1.015258304 | 0.021846827  | 1           |
| Lfng    | 3190.109042 | 3637.367867 | 1.140201736 | 0.189289103  | 1           |
| Mbp     | 3067.893197 | 3542.4208   | 1.154675399 | 0.207487339  | 1           |
| Smarca4 | 3015.518    | 3659.512139 | 1.213560038 | 0.279245484  | 0.544409111 |
| Ifnar1  | 2955.861012 | 4214.097712 | 1.425675191 | 0.511645332  | 0.027164565 |
| Hivep2  | 2941.717433 | 4241.903001 | 1.441981801 | 0.528052956  | 0.280718979 |
| Dicer1  | 2817.305704 | 3427.005911 | 1.216412513 | 0.282632562  | 0.683586066 |
| Il21r   | 2720.254218 | 2973.376012 | 1.093050786 | 0.128360435  | 1           |
| Il27ra  | 2691.442828 | 2786.775567 | 1.035420681 | 0.05021704   | 1           |
| Il6ra   | 2586.588247 | 3919.314056 | 1.515244671 | 0.599550768  | 0.673489052 |
| Ndfip1  | 2501.019294 | 2812.761211 | 1.124645946 | 0.169470893  | 1           |
| Foxo1   | 2453.241838 | 4482.327806 | 1.827103931 | 0.869558701  | 0.856657741 |
| Sit1    | 2398.131317 | 1939.690887 | 0.808834309 | -0.3060839   | 0.713721321 |
| Stat4   | 2355.275906 | 2391.405492 | 1.015339853 | 0.021962705  | 1           |
| Card11  | 2260.740093 | 3157.691698 | 1.396751315 | 0.482075179  | 0.681915016 |
| Rc3h1   | 2256.725285 | 3079.636435 | 1.364648349 | 0.448529237  | 0.140305445 |
| Notch1  | 2247.853209 | 3587.892135 | 1.596141652 | 0.674588692  | 0.002123265 |
| Ifngr1  | 2081.429128 | 3296.183741 | 1.583615649 | 0.663222229  | 0.000632313 |
| Smarce1 | 1879.242503 | 1709.098877 | 0.909461591 | -0.136915384 | 1           |
| Il18r1  | 1837.757503 | 2862.014478 | 1.557340658 | 0.639084559  | 0.133590579 |
| Foxo3   | 1795.445389 | 2361.880287 | 1.315484337 | 0.395594071  | 0.27774765  |
| Socs3   | 1776.943574 | 2932.247993 | 1.650163818 | 0.722609253  | 0.001292169 |
| Arrb1   | 1742.678526 | 1267.243847 | 0.727181651 | -0.459612299 | 0.088512575 |
| Cblb    | 1686.597023 | 2426.467776 | 1.438676662 | 0.524742387  | 0.545569375 |
| Mapk14  | 1614.207093 | 1594.481771 | 0.987780179 | -0.017738076 | 1           |
| Tnfsf8  | 1555.778106 | 2070.74613  | 1.331003516 | 0.412514382  | 0.085797259 |
| Runx1   | 1545.90878  | 2523.279876 | 1.632230768 | 0.706845043  | 0.119244645 |
| Prelid1 | 1518.113763 | 1571.213216 | 1.034977255 | 0.049599063  | 1           |
| Notch2  | 1514.651019 | 2087.599457 | 1.378270922 | 0.462859502  | 0.106528074 |
| Smad3   | 1475.885624 | 1796.562821 | 1.21727781  | 0.283658461  | 0.66462736  |
| Cd40lg  | 1443.933901 | 1421.720202 | 0.984615848 | -0.022367134 | 1           |
| Smad2   | 1402.595663 | 1415.266902 | 1.009034135 | 0.012974981  | 1           |
| Tgfb1   | 1380.13267  | 1663.387067 | 1.205237079 | 0.269316963  | 0.78001934  |
| Sh2d1a  | 1313.467883 | 883.2879018 | 0.672485344 | -0.572425268 | 0.033066041 |
| Mapk9   | 1174.101301 | 1388.634271 | 1.182721006 | 0.242109794  | 1           |
| Akt1    | 1156.305862 | 1192.258524 | 1.031092692 | 0.044174032  | 1           |
| Traf6   | 1152.136981 | 1406.08226  | 1.220412402 | 0.287368746  | 0.637027267 |
| Nfatc1  | 1125.331732 | 1676.130824 | 1.489454865 | 0.574784407  | 0.686766295 |
| Mtor    | 1114.853824 | 1360.127902 | 1.220005593 | 0.286887762  | 0.781239841 |
| Slc3a2  | 1106.912635 | 1124.630121 | 1.016006219 | 0.022909233  | 1           |
| Pik3cg  | 1074.009085 | 1508.626073 | 1.404667888 | 0.490229068  | 0.073274302 |
| Sema4a  | 1054.689372 | 1896.257225 | 1.797929585 | 0.84633652   | 0.000225857 |

|         |             |             |             |              |                  |
|---------|-------------|-------------|-------------|--------------|------------------|
| Abl1    | 1031.705978 | 1365.937737 | 1.323960282 | 0.404859843  | 0.310858919      |
| Irak1   | 993.6620453 | 1163.417286 | 1.170838004 | 0.22754148   | 1                |
| Mapk8   | 962.3893417 | 1127.364604 | 1.171422579 | 0.228261607  | 1                |
| Batf    | 939.2232927 | 975.2713912 | 1.038380754 | 0.054335549  | 1                |
| Stat5a  | 936.5972801 | 1376.710257 | 1.469906315 | 0.555724207  | 0.025518591      |
| Ifngr2  | 898.7976573 | 1111.140061 | 1.236251621 | 0.305972413  | 0.604340221      |
| Tyk2    | 897.3845628 | 958.4567785 | 1.06805579  | 0.094987009  | 1                |
| Cd274   | 799.7447617 | 1518.611369 | 1.898870042 | 0.925141171  | 0.000000485      |
| Nfatc2  | 750.8688575 | 1509.108448 | 2.009816273 | 1.007063624  | 0.005653796      |
| Jak3    | 746.1953627 | 1489.427514 | 1.996028906 | 0.997132614  | 0.0000131        |
| Rela    | 734.2516773 | 964.3891356 | 1.313431301 | 0.393340742  | 0.414430309      |
| Usp18   | 722.805453  | 1934.275099 | 2.676066003 | 1.420113699  | 0.058019726      |
| Icos    | 701.0018782 | 628.2454701 | 0.896210823 | -0.158089946 | 1                |
| Gata3   | 603.3615676 | 1081.575621 | 1.792582888 | 0.842039829  | 0.000398736      |
| Id3     | 602.8529025 | 322.349245  | 0.5347063   | -0.90318142  | 0.002395469      |
| Relb    | 455.9583937 | 962.0444057 | 2.109939019 | 1.077201303  | 0.0000000867     |
| Ptger4  | 451.8221389 | 600.4110575 | 1.328865954 | 0.410195583  | 0.460699495      |
| Icam1   | 408.3645802 | 633.0070503 | 1.550102729 | 0.63236383   | 0.002008982      |
| Ripk2   | 404.4509513 | 572.4440046 | 1.415360757 | 0.501169824  | 0.109206099      |
| Atp7a   | 401.2583844 | 470.7830617 | 1.173266603 | 0.230530877  | 1                |
| Runx3   | 396.3783316 | 679.1243892 | 1.713323699 | 0.776797746  | 0.038361505      |
| Egr2    | 392.8369464 | 336.2998817 | 0.856080073 | -0.224182351 | 1                |
| Hras    | 373.4753228 | 389.9094996 | 1.044003381 | 0.062126384  | 1                |
| P2rx7   | 350.556795  | 492.1276157 | 1.403845604 | 0.489384276  | 0.44077209       |
| Tnf     | 346.5566542 | 253.1443499 | 0.730455892 | -0.453130935 | 0.143138283      |
| Klf10   | 339.0194147 | 596.9246284 | 1.760738773 | 0.816180884  | 0.000434765      |
| Rel     | 330.4821765 | 769.4785775 | 2.328351216 | 1.219308695  | 0.0000346        |
| Fancc   | 299.4844313 | 314.280524  | 1.049405215 | 0.069571864  | 1                |
| Furin   | 290.2408711 | 451.6086837 | 1.555978943 | 0.637822536  | 0.152253146      |
| Zfp354c | 278.0152689 | 372.1283723 | 1.33851775  | 0.420636271  | 1                |
| Tlr7    | 270.2423198 | 802.8296949 | 2.970777099 | 1.570840362  | 0.000250577      |
| Bcl6    | 245.319328  | 190.908057  | 0.778202266 | -0.361782913 | 1                |
| Irf4    | 230.6154964 | 540.7704253 | 2.34490064  | 1.229526793  | 0.00000000015    |
| Tgfb1   | 228.5424794 | 616.7922857 | 2.698808062 | 1.432322376  | 0.00000000000137 |
| Rora    | 226.7049103 | 291.9506449 | 1.287800271 | 0.364908859  | 1                |
| Bcl3    | 217.6967083 | 458.9973841 | 2.108425928 | 1.076166338  | 0.00000186       |
| Cd86    | 182.8432662 | 277.143221  | 1.515742016 | 0.600024224  | 1                |
| Rsad2   | 174.8313897 | 804.0795562 | 4.599171565 | 2.201374016  | 0.130576871      |
| Prdm1   | 171.6955146 | 102.8518129 | 0.599036108 | -0.739285127 | 0.019427249      |
| Maf     | 164.5489141 | 116.1954795 | 0.706145526 | -0.501962563 | 0.262811606      |
| Ctla4   | 161.4371095 | 285.5085384 | 1.768543424 | 0.822561643  | 0.000138353      |
| Cebpb   | 161.2369365 | 86.46235698 | 0.536244107 | -0.899038208 | 0.054676134      |
| Gata1   | 138.0911139 | 328.1468967 | 2.376307117 | 1.248721304  | 0.0000723        |

|          |             |             |             |              |              |
|----------|-------------|-------------|-------------|--------------|--------------|
| Il2rb    | 122.4887043 | 547.6536471 | 4.471054291 | 2.160615064  | 3.73E-37     |
| Gadd45b  | 98.80518435 | 131.556918  | 1.331477886 | 0.413028468  | 0.81296247   |
| Cish     | 89.25944152 | 436.7543116 | 4.893088105 | 2.29074526   | 1.38E-33     |
| Klf4     | 82.14295336 | 109.4716297 | 1.33269654  | 0.414348311  | 1            |
| Eomes    | 74.97614543 | 12.15627632 | 0.162135253 | -2.624730284 | 0.000442713  |
| Gadd45g  | 61.94341286 | 92.73070065 | 1.497022789 | 0.582096183  | 0.247399283  |
| Nr4a2    | 39.98073551 | 50.4999466  | 1.263106993 | 0.336976849  | 1            |
| Foxp3    | 38.36597701 | 179.645722  | 4.682422709 | 2.22725518   | 0.0000126    |
| Efnb1    | 34.71236339 | 49.37105815 | 1.422290312 | 0.508215972  | 1            |
| Ahr      | 32.29622607 | 68.96646531 | 2.135434189 | 1.094529437  | 0.026783791  |
| Il2ra    | 26.38025084 | 157.2582893 | 5.961212811 | 2.575605878  | 3.17E-30     |
| Il12rb1  | 22.35821676 | 78.54542537 | 3.513045169 | 1.812722126  | 0.0000000829 |
| Il12rb2  | 17.72083539 | 20.22325315 | 1.141213307 | 0.190568474  | 1            |
| Pcgf2    | 13.82075405 | 4.254681203 | 0.307847256 | -1.699713388 | 0.142920485  |
| Bhlhe41  | 8.957472368 | 8.178284683 | 0.913012549 | -0.131293405 | 1            |
| Il1rl1   | 8.037098854 | 2.534310151 | 0.315326487 | -1.665081733 | 0.811513738  |
| Il21     | 7.324727008 | 2.775740838 | 0.378954852 | -1.399902117 | 0.538068849  |
| Il2      | 5.9149617   | 3.540033993 | 0.598488067 | -0.740605613 | 1            |
| Il1r1    | 5.100109944 | 7.523876466 | 1.475238093 | 0.560947814  | 1            |
| Rorc     | 5.039615124 | 3.704781801 | 0.735131892 | -0.443924984 | 1            |
| Notch3   | 4.979120305 | 2.114533589 | 0.424680156 | -1.235551397 | 1            |
| Pparg    | 4.586967824 | 0           | 0           | #NOM?        | 0.076463124  |
| Tbx21    | 3.883745943 | 0.825955058 | 0.212669693 | -2.233313641 | 0.883399751  |
| Anxa1    | 3.615203977 | 7.034216057 | 1.945731445 | 0.960312599  | 0.859069513  |
| Il23r    | 3.027731139 | 0.925246946 | 0.305590855 | -1.710326727 | 1            |
| Il10     | 2.323731322 | 0           | 0           | #NOM?        | 0.452188933  |
| Ccl3     | 2.263236502 | 0           | 0           | #NOM?        | 0.97479253   |
| Cd80     | 1.810589202 | 1.651910116 | 0.912360526 | -0.132324067 | 1            |
| Il17f    | 1.312525995 | 0.825955058 | 0.629286628 | -0.668210807 | 1            |
| Il5      | 0.905294601 | 0           | 0           | #NOM?        | 1            |
| Hlx      | 0.905294601 | 1.338224475 | 1.478219879 | 0.56386088   | 1            |
| Ifng     | 0.4526473   | 4.477296898 | 9.891358888 | 3.306168734  | 0.594701315  |
| Tnfrsf8  | 0.4526473   | 4.229067177 | 9.342963436 | 3.223880222  | 0.08692993   |
| Gfi1b    | 0.23618402  | 0           | 0           | #NOM?        | 1            |
| Il4      | 0           | 0.462623473 | Inf         | Inf          | 1            |
| Il13     | 0           | 0.462623473 | Inf         | Inf          | 1            |
| Il17a    | 0           | 1.387870419 | Inf         | Inf          | 1            |
| Tnfsf13b | 0           | 0           | NA          | NA           | NA           |

**Table S2.** Related to Figure 3. mRNA levels of genes related to the Th1- or Th2-lineages in Setdb1<sup>+/+</sup> and Setdb1<sup>-/-</sup> Th2 cells following 2 days of culture in Th1-polarizing conditions, as measured by RNA-seq.



## Th2 gene set

| gene_name | baseMean Setdb1+/+ | baseMean Setdb1-/- | foldChange         | log2FoldChange       | padj                  |
|-----------|--------------------|--------------------|--------------------|----------------------|-----------------------|
| Itgb2     | 4403.59209862879   | 6640.71306272036   | 1.50802184080314   | 0.592657323427667    | 0.0127479449600569    |
| Pparg     | 3071.54714714301   | 633.448114547234   | 0.206230959253363  | -2.27766716951109    | 2.59541229871786E-11  |
| Il4       | 37334.3170973556   | 3100.8983199724    | 0.0830575877921186 | -3.58974421701133    | 1.68948286128327E-72  |
| Stat6     | 2851.72665792246   | 3290.16261023437   | 1.15374403121487   | 0.206323184244631    | 0.640680497254622     |
| Stat5a    | 1579.73542970862   | 2778.53075934991   | 1.75885829177257   | 0.814639252001168    | 0.000782665335139924  |
| Csf1      | 4284.61880269507   | 1251.35110167995   | 0.292056577096855  | -1.77568022033511    | 0.0000527618968298106 |
| Gzmb      | 3670.12463114639   | 21641.7856690862   | 5.89674407387256   | 2.55991858073484     | 1.82027481233416E-11  |
| Hivep2    | 1213.36981692906   | 1889.13068633912   | 1.55692902524999   | 0.638703178679159    | 0.0973856771383928    |
| Gata3     | 3086.89142875051   | 1301.29968769488   | 0.421556675293115  | -1.24620149060565    | 7.59135985050001E-07  |
| Il10      | 53035.4749289679   | 11739.7937952443   | 0.221357380337741  | -2.17555061935465    | 2.23466528154697E-16  |
| Nbr1      | 2757.77594152984   | 2827.9603777021    | 1.0254496513351    | 0.0362566585146333   | 1                     |
| Pcgf2     | 222.025256027969   | 95.238181451965    | 0.428952017242434  | -1.22111181862638    | 0.0403607687408139    |
| Tube1     | 497.589547146432   | 744.369178585987   | 1.49595019199013   | 0.581062141157983    | 0.0247913164220133    |
| Socs2     | 1239.77975323469   | 1684.72870757005   | 1.35889354796645   | 0.442432443716657    | 0.144426099029632     |
| Il13      | 186556.91930374    | 22908.1203406086   | 0.122794267969825  | -3.02568487753852    | 1.09055247353345E-09  |
| Itk       | 23481.5526867978   | 20487.1216445862   | 0.872477298151786  | -0.196810501983335   | 0.630957311022026     |
| Rsad2     | 205.849214287881   | 907.108016547496   | 4.40666251598562   | 2.13968641212306     | 0.211483557732428     |
| Itgb3     | 3296.03064627094   | 1596.21711723791   | 0.48428467103115   | -1.04607275657169    | 6.35499580476533E-07  |
| Stat5b    | 1749.86726706641   | 3232.67098496747   | 1.84738068184276   | 0.885481186957255    | 0.0000913003513193384 |
| Hif1a     | 66477.0224329862   | 58350.4797912015   | 0.877754111956853  | -0.188111245267103   | 0.655771623804097     |
| Asb2      | 2119.45742927579   | 3206.11241387544   | 1.51270432182742   | 0.597130021183016    | 0.0176779221570156    |
| Irf4      | 11626.3002419059   | 14513.2676187687   | 1.24831350617086   | 0.319980303589457    | 0.347809993277672     |
| Pros1     | 2037.85068089764   | 2304.44938410186   | 1.13082347283992   | 0.177373735043559    | 0.759656164795013     |
| Gzma      | 7937.11621679942   | 10687.7112812616   | 1.34654841750211   | 0.429266105394248    | 0.537755426754284     |
| Epas1     | 9277.55487024298   | 2411.54966010979   | 0.259933753433747  | -1.94378410921244    | 1.61027374104615E-09  |
| Anxa1     | 1276.39447871919   | 579.002260404697   | 0.453623288143414  | -1.14043338752922    | 0.00125428355376212   |
| Prdx4     | 2188.04354451669   | 1394.85762697487   | 0.637490798787085  | -0.649523575880638   | 0.00735707586362433   |
| Il1rl1    | 763.227941940993   | 266.524898225654   | 0.349207469459051  | -1.51784367670958    | 0.000809941370178045  |
| Il1r2     | 6364.43591139998   | 1156.39084718275   | 0.18169573286321   | -2.4604035567098     | 4.49268357348027E-19  |
| Rgs1      | 86615.5812157871   | 37557.3444147036   | 0.433609564093742  | -1.20553151642796    | 1.84630006741086E-09  |
| Il24      | 7222.67780828239   | 6387.1428947615    | 0.884317847798394  | -0.177363088366086   | 0.705777984113408     |
| Tnfrsf4   | 351.261551913896   | 690.93905095887    | 1.96702157464771   | 0.97601278144286     | 0.00727325705665794   |
| Il2ra     | 58753.1312543135   | 40015.866598456    | 0.681084833168244  | -0.554093589206976   | 0.0159720112233453    |
| Prkcg     | 2351.67298977384   | 2892.58543280416   | 1.23001175987583   | 0.298672108925029    | 0.349277745199831     |
| Notch1    | 3832.1024470119    | 6844.05037570809   | 1.78597792474071   | 0.836714248396463    | 0.000612572229210465  |
| Il2       | 174.63333542678    | 14.3063743223687   | 0.0819223562752548 | -3.60959897861482    | 0.063538877428868     |
| Ecm1      | 1369.87472556188   | 741.93376158404    | 0.541607015400422  | -0.884681668718454   | 0.118527440267545     |
| Nfkb1     | 10064.12944031     | 13348.1871698679   | 1.3263131450202    | 0.407421438671711    | 0.125707964139786     |
| Tnfrsf8   | 4609.19999763584   | 7309.36418209457   | 1.58582057316751   | 0.665229547306037    | 0.0146860895660884    |
| Tnfrsf9   | 36060.0885782616   | 17782.622052117    | 0.493138612611093  | -1.01993487502016    | 0.00200007495515737   |
| Gfi1      | 1729.28551560481   | 1657.38086918748   | 0.958419447934726  | -0.0612709117665945  | 1                     |
| Lfng      | 960.536157453348   | 1780.38918254958   | 1.85353686973106   | 0.890280812890613    | 0.000158231254602394  |
| Bhlhe41   | 376.38606784287    | 21.0561863685496   | 0.05594305466519   | -4.15989715816294    | 1.74096275491741E-26  |
| Il4ra     | 10856.4098957721   | 12788.522685887    | 1.17796977165234   | 0.236302518076012    | 0.514719181350501     |
| Rnf128    | 424.134495663478   | 134.196633209931   | 0.316401128844768  | -1.66017334787108    | 1.18153424145647E-11  |
| Dusp4     | 2646.69362406938   | 5969.12989451676   | 2.25531577974599   | 1.17332944766172     | 0.000033952636067909  |
| Jak3      | 2035.67638415405   | 3437.22197529555   | 1.688491354545     | 0.755734792441641    | 0.011236745583308     |
| Cyp11a1   | 11990.565348413    | 2213.92900361765   | 0.184639250884919  | -2.43721881905608    | 2.64103888591819E-33  |
| Nlrp3     | 409.477947001474   | 105.991412038992   | 0.258845226745777  | -1.94983838056602    | 9.97996090990892E-06  |
| Myo6      | 218.829811534824   | 210.902941904718   | 0.963776098080474  | -0.053230072610416   | 1                     |
| Vav1      | 7899.59390068863   | 5890.07721556005   | 0.745617722835929  | -0.423491942588802   | 0.0991640137776258    |
| Batf      | 13084.9812729995   | 3185.76701550861   | 0.243467449363673  | -2.03819919231565    | 4.52539832924973E-25  |
| Jdp2      | 1747.30120967634   | 388.945062351434   | 0.32297603777474   | -2.16749003247237    | 2.4657080990598E-12   |
| Egln3     | 7013.59311271106   | 4018.60445599984   | 0.572973708542735  | -0.803459153854993   | 0.07073315389874      |
| Tanc1     | 343.1026119893     | 308.243405095144   | 0.89840005387297   | -0.154570080636978   | 1                     |
| Tmtc2     | 582.156438158222   | 610.537167697966   | 1.04875103611245   | 0.0686722359747994   | 1                     |
| Il5       | 156124.070368313   | 28808.127237977    | 0.28850278738468   | -2.4381451996403     | 4.54166840843288E-07  |
| Prdm1     | 1889.14342160676   | 2467.54998121264   | 1.3061739796939    | 0.385347073702523    | 0.35176955362482      |
| Rbpj      | 6615.29026951315   | 7881.32281358598   | 1.19137974185462   | 0.252633332907939    | 0.488539641502186     |
| Stx11     | 8296.69262669857   | 4446.66373536029   | 0.535956185848205  | -0.899813028986325   | 0.000167012477318073  |
| Ccr8      | 1380.9671614996    | 796.72272202073    | 0.57693096855074   | -0.793529388299915   | 0.0777988252761089    |
| Ptgir     | 714.19286413565    | 691.24501309351    | 0.9678688318037    | -0.0471165520658105  | 1                     |
| S100a1    | 339.548201060576   | 361.452303274087   | 1.06450955164861   | 0.0901888948858025   | 1                     |
| Ctla2a    | 19439.1701833387   | 6122.92498291648   | 0.314978722094035  | -1.6666737220407     | 1.55515406904202E-07  |
| Zfp354c   | 26.4499730959328   | 14.3094568408571   | 0.5410080892171    | -0.886297343669897   | 0.422954778832917     |
| Ccr4      | 846.779833243622   | 2185.24309420989   | 2.58065084738643   | 1.36773496327677     | 1.77919824195186E-08  |
| F2r       | 6334.59987658586   | 4690.27814860573   | 0.740422163985775  | -0.433580013031425   | 0.0952706243221149    |
| Rictor    | 1528.21190887766   | 2377.61429417671   | 1.55581453093299   | 0.637670086512129    | 0.00917919912350192   |
| Cysltr1   | 2450.76661728116   | 2440.40804820459   | 0.995773335166422  | -0.00611071147599898 | 1                     |
| Junb      | 6552.40512168505   | 9401.25660467386   | 1.43477950921572   | 0.520829046725364    | 0.396368624736714     |
| Socs3     | 2216.17069863577   | 1369.1464329837    | 0.617798274215302  | -0.694792254003558   | 0.00890949407781522   |
| Bcl3      | 467.992638535135   | 679.692609655512   | 1.45235748105573   | 0.5383965998005      | 0.125110602989016     |
| Tanc2     | 3623.96128482568   | 1708.86581574288   | 0.471546377412549  | -1.08452842542346    | 0.00726502203041791   |
| Maf       | 4093.71104289548   | 1207.20770686922   | 0.294893238486946  | -1.76173535120315    | 1.32314300088593E-15  |
| Nfil3     | 3269.52369200975   | 3974.45516931193   | 1.2156067805916    | 0.281676627273431    | 0.550849521451797     |
| Crem      | 6173.46379278666   | 3589.5788664412    | 0.581452971447798  | -0.782265585222871   | 0.000382536315874869  |
| Cst7      | 4618.50361436699   | 2795.25689945865   | 0.6052299067562    | -0.7244446173676058  | 0.00159093801040407   |
| Il2rb     | 7302.11394919119   | 14445.7535159845   | 1.97829746515864   | 0.984259372124867    | 5.31863742817592E-06  |
| Ctla2b    | 1254.13800545061   | 463.244658802134   | 0.369372953206765  | -1.43684986392986    | 6.79356636732509E-09  |



## Th1 gene set

| gene_name | baseMean Setdb1+/+ | baseMean Setdb1-/- | foldChange         | log2FoldChange      | padj                  |
|-----------|--------------------|--------------------|--------------------|---------------------|-----------------------|
| Mx1       | 1043.35931955026   | 3939.02293451553   | 3.77532731122145   | 1.91660172782996    | 0.00469743657838693   |
| Il12rb1   | 2765.89129090897   | 5182.4040854136    | 1.87368321468284   | 0.90587705585799    | 0.0140186173169832    |
| Tbx21     | 168.428010164137   | 2170.80314791642   | 12.8886112577173   | 3.68802491719513    | 1.37201404327399E-12  |
| Napsa     | 62.1357854231987   | 201.492077091266   | 3.24277026063049   | 1.69722681732142    | 0.255109537000213     |
| Relb      | 728.473501987121   | 998.919826231886   | 1.3712507366528    | 0.455492395662239   | 0.165686195110904     |
| Stat5a    | 1579.73542970862   | 2778.53075934991   | 1.75885829177257   | 0.814639252001168   | 0.00782665335139924   |
| Il27ra    | 752.105171008521   | 446.607680839998   | 0.593810145250202  | -0.751926352898055  | 0.00233673647910785   |
| Prg4      | 2.36231968339229   | 2.95104428034594   | 1.24921461777275   | 0.32102135636743    | 1                     |
| Slc3a2    | 2010.01949907757   | 4343.03884349593   | 2.16069488156161   | 1.11149535913079    | 7.00429017543593E-07  |
| Mcoln2    | 47.6138322909016   | 242.417830142393   | 5.0913320453039    | 2.34804315801815    | 3.45737545556964E-12  |
| Gadd45b   | 1361.33003796211   | 1771.43585822816   | 1.30125378036906   | 0.37990235481136    | 0.535417542106688     |
| Hivep2    | 1213.36981692906   | 1889.13068633912   | 1.55692902524999   | 0.638703178679159   | 0.0973856771383928    |
| Cd274     | 3161.04610405251   | 4540.47940940287   | 1.43638506366038   | 0.522442556238901   | 0.0438080814250292    |
| Mfng      | 155.842727318839   | 333.517758827672   | 2.14009190268678   | 1.09767275209537    | 0.00246067522812143   |
| Il12rb2   | 466.365196097854   | 6491.79068257564   | 13.9199724526903   | 3.79908445101607    | 3.86785458206608E-11  |
| Irf1      | 1705.46705325169   | 7403.81181476233   | 4.34122242387857   | 2.11810134141272    | 1.41130949576195E-23  |
| Ccl9      | 715.503701074735   | 623.315239889256   | 0.87115585698997   | -0.198997242893088  | 1                     |
| Ifngr1    | 608.61734265779    | 2592.66122667904   | 4.25992005971611   | 2.09082635755599    | 4.32924486350282E-21  |
| Rel       | 936.072593230275   | 847.450434910839   | 0.905325549577718  | -0.143491425005116  | 0.971425532455332     |
| Mapk9     | 728.287938263474   | 909.120058608184   | 1.24829756315323   | 0.319961877882103   | 0.395436641467078     |
| Stat5b    | 1749.86726706641   | 3232.67098496747   | 1.84738068184276   | 0.885481186957255   | 0.0030913003513193384 |
| Gadd45g   | 629.256075284645   | 3816.92351779195   | 6.06577142074529   | 2.60069113324554    | 5.42273906056877E-22  |
| Cd86      | 0.590579920848073  | 15.0848969915598   | 25.542515854413    | 4.67482872767768    | 0.00365645864269464   |
| Ifngr2    | 404.685395065796   | 245.749594368284   | 0.607260843520998  | -0.719611748388847  | 0.011978789874432     |
| Ifnar1    | 1674.74965478819   | 2021.22531103284   | 1.20688205861339   | 0.271284696998198   | 0.455145996282436     |
| Slc11a2   | 195.013808397856   | 343.757340170297   | 1.7627333315238    | 0.817814238477001   | 0.023958104280182     |
| Tnf       | 1451.32856980821   | 1342.38963781681   | 0.924938477573138  | -0.112570687152414  | 0.965951893800375     |
| Anxa1     | 1276.39447871919   | 579.002260404697   | 0.453623288143414  | -1.14043338752922   | 0.00125428355376212   |
| Jak2      | 20839.944033208    | 19362.018929909    | 0.929082098255928  | -0.106122009029222  | 0.93377733281912      |
| Rela      | 923.484851168819   | 1186.88005709101   | 1.2852187619416    | 0.362013946894984   | 0.342522142712275     |
| Hras      | 1053.36146556146   | 1035.91508287754   | 0.983437420814883  | -0.0240948426402459 | 1                     |
| Il18rap   | 512.817272138172   | 594.624758710473   | 1.15952560691103   | 0.213534680747721   | 0.863123279595794     |
| Il18r1    | 238.077633641985   | 383.244584400773   | 1.6097462770362    | 0.68683331336628    | 0.167961510577557     |
| Stat1     | 8616.07542910562   | 20418.477473816    | 2.36981182927451   | 1.24477250901799    | 0.0035698065575694    |
| Slc11a1   | 8.74227726604683   | 7.57329100139835   | 0.866283551862559  | -0.207088769809714  | 1                     |
| Tnfrsf4   | 351.261551913896   | 690.93905095887    | 1.96702157464771   | 0.97601278144286    | 0.00727325705665794   |
| Il2ra     | 58753.1312543135   | 40015.866598456    | 0.681084833168244  | -0.554093589206976  | 0.0159720112233453    |
| Abl1      | 566.027675787836   | 845.723166046655   | 1.49413748165144   | 0.57931290245864    | 0.0426144465595361    |
| Il2       | 174.63333542678    | 14.3063743223687   | 0.0819232562752548 | -3.60959897861482   | 0.063538877428868     |
| Sema4a    | 113.660535862244   | 460.662277646395   | 4.05296591426169   | 2.01897804226548    | 1.41511626741734E-12  |
| Jak1      | 4586.92521531515   | 5362.03640843068   | 1.16898274044834   | 0.225253629225074   | 0.558593770702632     |
| Mtor      | 1314.67668343127   | 1869.34611406913   | 1.42190558152305   | 0.507825669183091   | 0.0722085953206278    |
| P2rx7     | 42.8697893330046   | 791.859080591832   | 18.4712612987392   | 4.20721047811661    | 4.3680612273994E-30   |
| Lfng      | 960.536157453348   | 1780.38918254958   | 1.85353686973106   | 0.890280812890613   | 0.00158231254602394   |
| Furin     | 10376.8917942544   | 14343.9688415267   | 1.38229916297949   | 0.467069883627128   | 0.303267660392952     |
| Ets1      | 16055.6657822269   | 15853.9734584132   | 0.987437934586495  | -0.0182380244604163 | 1                     |
| Mplz12    | 6.46727374266023   | 72.0586475393847   | 11.1420438358844   | 3.47794199196727    | 0.0030669323815881672 |
| Tyk2      | 1254.84377731005   | 1444.53590574984   | 1.1511679237446    | 0.203098298355443   | 0.717653236161573     |
| Eomes     | 4.11465585482413   | 24.7053730776054   | 6.00423800902818   | 2.58598116677187    | 0.0132096880016162    |
| Cxcl10    | 10.6013331885967   | 81.2569378333261   | 7.66478483298022   | 2.93824529289521    | 0.00169180700034303   |
| Dnajc12   | 167.615518553576   | 117.928768596947   | 0.703567125613447  | -0.507240020864401  | 0.685612726009554     |
| Prkcdp    | 161.447770865078   | 9.75055761389087   | 0.0603945013402471 | -4.04943898542758   | 1.54564558368183E-09  |
| Socs1     | 692.294127694569   | 1786.69370723337   | 2.58083036639831   | 1.36783531865056    | 3.09399511699572E-06  |
| Notch3    | 0.580878125291885  | 5.59593016565575   | 9.63357014493162   | 3.2680705515343     | 0.152867234839091     |
| Prdm1     | 1889.14342160676   | 2467.54998121264   | 1.3061739796939    | 0.385347073702523   | 0.35176955362482      |
| Hlx       | 22.8288792063949   | 204.651693612423   | 8.96459663052995   | 3.16423866929261    | 0.0030376975302637742 |
| Rhou      | 0.580878125291885  | 22.3969512486327   | 38.5570574505255   | 5.26892304885357    | 0.0871827580455445    |
| Xaf1      | 1683.63390523242   | 1541.74617967081   | 0.915725309925958  | -0.127013196826941  | 1                     |
| Cxcr5     | 7.5514156287945    | 2.46108540424013   | 0.325910468343935  | -1.61745240226063   | 0.511113945843129     |
| Myof      | 54.3612284966119   | 15.3578827092049   | 0.282515372333098  | -1.82359872475653   | 0.215625725102827     |
| Foxo3     | 754.642123011925   | 1433.32792087777   | 1.89934788579926   | 0.925504174655947   | 0.000093318684450115  |
| Cxcr3     | 8.14199554964257   | 18.4614173913162   | 2.26743152569355   | 1.18105898337388    | 0.811562895813183     |
| Socs3     | 2216.17069863577   | 1369.1464329837    | 0.617798274215302  | -0.694792254003558  | 0.00390949407781522   |
| Uck2      | 26.3820605270395   | 88.2469015762613   | 3.34495865043655   | 1.74198837673255    | 0.00257003270597034   |
| Ifng      | 57.2850227141836   | 21293.1063853379   | 212.849812325594   | 7.73369201393885    | 0.0022204460012758    |
| Ms4a4b    | 0.590579920848073  | 23.8540475469685   | 40.3908881844714   | 5.33595796574251    | 2.30799277766536E-08  |
| Stat4     | 625.434932512665   | 1255.99429872687   | 2.00819339220621   | 1.00589820979768    | 0.0000188175242596769 |
| Il2rb     | 7302.11394919119   | 14445.7535159845   | 1.97829746515864   | 0.984259372124867   | 5.31863742817592E-06  |
| Runx3     | 375.437751094522   | 692.848273137706   | 1.84544114468465   | 0.883965727491103   | 0.023472618266088     |
| Cd80      | 31.5129456991046   | 56.483202874428    | 1.79238092857923   | 0.841877281016946   | 0.664841658863554     |

**Table S3.** Related to Figure 4. Family, name and coordinates of H3K9me3+ ERVs in Setdb1<sup>+/+</sup> Th2 cells, as determined by ChIP-seq.

#### ACKNOWLEDGMENTS

We greatly acknowledge T. Jenuwein (Max Planck Institute, Freiburg, Germany) for providing the Suv39h1-deficient mice. We also thank F.-E. L'Faqihi-Olive, V. Duplan-Eche and A.-L. Iscache for technical assistance at the flow cytometry facility of INSERM U1043, the personnel of the US006 ANEXPLO/CREFRE animal facility for expert animal care, D. Rozet for administrative assistance, the GeT and Bioinformatics platforms from the Genotoul (Toulouse, Région Occitanie, France), the Transcriptomic & Genomic Platform Marseille Luminy (TGML, Marseille, France), and the Genomic and Transcriptomic platform from the CPTP (Toulouse, Région Occitanie, France). We are also grateful to members of the "T cell-mediated Immune Tolerance" laboratory for advice, discussions and technical help. We would like to thank D. Dunia for his careful reading of the manuscript and his comments. This work was supported by Agence Nationale de la Recherche (ANR JCJC "EpiTreg" to O.J.), Région Occitanie (NVEQ 2014 to O.J.) and Fondation pour la Recherche Médicale (FRM, AJE201212 to O.J.). B.B. was funded by a fellowship from the Association pour la Recherche sur le Cancer (ARC). A.M. was supported by fellowships from the Région Midi-Pyrénées and FRM.

#### AUTHOR CONTRIBUTIONS

V.A., B.B. and A.M. performed the main experiments, analyzed data and contribute to writing the manuscript. J.F. assisted V.A. with ChIP-Seq and RNA-Seq data analyses. V.A., P.R., J.P.M.v.M. and S.A supervised the study. O.J conceived and supervised the study and wrote the manuscript.

#### CORRESPONDING AUTHORS

Correspondence should be addressed to Olivier P Joffre (olivier.joffre@inserm.fr) or Véronique Adoue (veronique.adoue@inserm.fr).

#### COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.





## DISCUSSION ET PERSPECTIVES

Les lymphocytes T CD4 orchestrent et coordonnent les réponses immunitaires. Leur atout repose essentiellement sur leur capacité à adapter très finement leur phénotype et leurs fonctions au danger détecté par les cellules de l'immunité innée et les tissus. Ces cellules émettent en effet des signaux qualitativement différents selon la menace qu'elles détectent, ce qui permet au lymphocyte T CD4 naïf, très plastique, de mettre en place un programme d'expression génique qui lui permettra, de façon directe ou indirecte, d'éliminer le danger tout en préservant l'intégrité des tissus. Le processus de différenciation des lymphocytes T CD4, quel que soit le lignage choisi, nécessite un remodelage massif de son paysage épigénétique. La chromatine va ainsi s'ouvrir au niveau des *loci* spécifiques du lignage ou se fermer au niveau des régions spécifiques des phénotypes alternatifs. L'épigénome contrôle donc l'acquisition de fonctions des cellules T en réponse à leur environnement mais également la balance plasticité / stabilité. Sa manipulation est donc d'un intérêt majeur en immunothérapie. Moduler les mécanismes épigénétiques qui contrôlent la biologie des lymphocytes T CD4 permettrait par exemple de contrôler l'amplitude de la réponse lymphocytaire, de la stabiliser dans un phénotype donné, ou de la détourner d'un type de réponse vers un autre. De nombreux processus épigénétiques ont été décrits tant dans l'ouverture de la chromatine que dans sa répression. Cependant peu de travaux ont étudié le rôle de l'interactome d'H3K9me3 dans la biologie des lymphocytes T CD4. C'est ce que j'ai réalisé durant ma thèse.

Je me suis plus particulièrement intéressée à l'impact de la lysine méthyltransférase SETDB1 sur l'homéostasie et la différenciation des lymphocytes T CD4. Nous avons montré *in vitro* et *in vivo* que SETDB1 réprime la programmation Th1. Au niveau cellulaire, ceci se traduit par une réponse Th1 exacerbée et par une instabilité des cellules Th2. D'un point de vue mécanistique, nous avons mis en évidence que SETDB1 dépose H3K9me3 sur un set restreint de rétrovirus endogènes qui se comportent comme des modules cis-régulateurs du programme d'expression Th1. SETDB1 est ainsi un garant de l'intégrité cellulaire et de la « barrière de lignage » existant entre les lymphocytes de type Th1 et Th2.

## I. Coopération des mécanismes épigénétiques dans la répression de lignages alternatifs de lymphocytes T helper

La déposition d'H3K9me3 par SETDB1 sur les ERV régulant le programme Th1 dans des lymphocytes Th2 s'ajoute à la liste des mécanismes épigénétiques déjà décrits comme contrôlant la détermination des cellules Th2. (**Figure 23**)

1) Le groupe du Dr. Amigorena a mis en évidence que la déposition d'H3K9me3 par SUV39H1 sur le promoteur de l'*Ifng* est nécessaire à la répression de ce locus dans les lymphocytes Th2 (212). Si, en l'absence de SUV39H1, les auteurs ont observé un défaut de stabilité des cellules Th2 mutantes, le phénotype est dû à une perte seulement partielle d'H3K9me3 sur le promoteur de l'*Ifng*. Même si SUV39H1 doit agir sur d'autres *loci*, puisqu'une simple dérégulation du gène codant pour l'IFN- $\gamma$  ne pourrait expliquer à elle seule la déstabilisation des cellules Th2, ce mécanisme ne peut pour le moment être généralisé car les auteurs n'ont pas observé de diminution d'H3K9me3 sur *Tbx21* dans les LT Th2 mutants. De plus, la perte partielle de signal d'H3K9me3 observée dans les cellules déficientes pour SUV39H1 suggère qu'une autre KMT est responsable de la déposition d'H3K9me3 sur ces *loci* ou qu'une autre KMT compense l'absence de l'enzyme.

2) Le groupe du Dr. Wells a mis en évidence un rôle clef pour DNMT3A (269). Dans les lymphocytes Th2 déficients pour cette enzyme, la stabilité du lignage est altérée par l'absence de méthylation de l'ADN au niveau du promoteur de l'*Ifng*. Ainsi, lorsqu'ils sont cultivés en condition pro-Th1, les lymphocytes produisent simultanément de l'IFN- $\gamma$  et de l'IL-13. Les auteurs n'ont cependant étudié les marques épigénétiques que sur le locus de l'*Ifng*.

3) Enfin le groupe du Dr. Motohashi a montré que la déposition d'H3K27me3 par EZH2 permettait de réprimer le programme Th1 dans les LT Th1 et Th2 (270). Ainsi, de façon similaire à ce que nous observons dans les lymphocytes *Setdb1*<sup>-/-</sup>, EZH2 permet de freiner la programmation Th1 dans les LT Th1 et de la verrouiller dans les cellules Th2, assurant leur stabilité phénotypique et fonctionnelle. La perte de la marque dans les lymphocytes T déficients pour EZH2 a été observée sur les régions enhancers et les promoteurs de l'*Ifng* et de *Tbx21*.

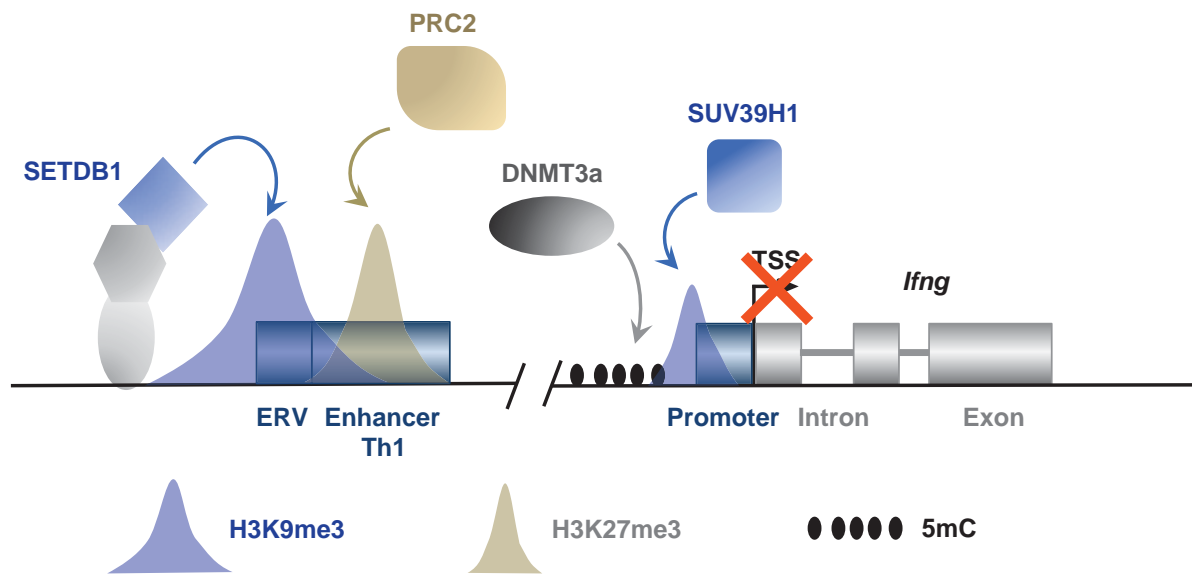


Figure 26 : Répression épigénétique du locus *Ifng* dans les lymphocytes Th2

Plusieurs voies de répression épigénétique, non redondantes et potentiellement interdépendantes, se coordonnent pour verrouiller le locus *Ifng* dans les lymphocytes Th2. SETDB1 cible la déposition d'H3K9me3 sur des ERVs se comportant comme des modules *cis*-régulateurs de gènes Th1. SUV39H1 est impliquée dans la déposition d'H3K9me3 au niveau du promoteur de l'*Ifng*. PRC2 via la déposition d'H3K27me3 sur les enhancers, et DNMT3a via la méthylation des cytosines, répriment également l'expression de l'*Ifng* dans des lymphocytes Th2.

Les trois études citées précédemment se sont focalisées sur quelques loci pro-Th1 (i.e. *Ifng* et *Tbx21*). On peut donc s'interroger sur leur impact à l'échelle du génome mais également sur la nature des régions qu'elles ciblent. En effet, si l'action de SETDB1 semble essentiellement cibler les enhancers pro-Th1, il serait intéressant de déterminer si les autres voies, en plus des promoteurs, pourraient aussi affecter d'autres éléments génomiques tels que ces régions *cis*-régulatrices distantes. De plus, le groupe du Dr. Wells a révélé que l'absence d'ADN méthylé sur le promoteur de l'*Ifng* s'accompagnait d'une diminution de l'accumulation d'H3K27me3. Cette observation suggère que ces voies sont interconnectées tout comme le sont, dans de nombreux systèmes la déposition d'H3K9me3 et la méthylation de l'ADN. Il serait donc intéressant d'étudier dans le temps et l'espace si, et comment, ces voies impactent les unes sur les autres à l'échelle du génome.

Si les voies de répression de la programmation Th1 dans les lymphocytes Th2 sont de mieux en mieux documentées, qu'en est-il du contrôle de la balance entre stabilité et plasticité dans les autres lignages ? La voie répressive DNMT3A/5mC a également été impliquée dans la répression de l'*Ifng* dans les lymphocytes Th17 et dans les lymphocytes T régulateurs (269).

Cette voie semble donc avoir un effet stéréotypé sur la répression du programme Th1 quel que soit le lignage concerné. La voie EZH2/H3K27me3 réprime la production des cytokines Th1 ou Th2 de façon indépendante du lignage dans lequel s'est engagé le lymphocyte (270). Elle semble donc fonctionner comme un frein global à la différenciation et permet sans doute d'éviter un emballement des réponses immunitaires qui pourrait être encore plus délétère pour l'organisme que le danger initial. Il est envisageable que la voie SETDB1/H3K9me3 soit impliquée dans la répression du programme Th1 dans les autres lignages de lymphocytes T helper, ou fonctionne comme un frein à leur différenciation. Néanmoins, une récente étude sur un de ses partenaires privilégiés, TRIM28, laisse imaginer que sa fonction ne sera pas aussi stéréotypée. En effet, il a été décrit que le recrutement de TRIM28 au niveau des enhancers et promoteurs des gènes codant pour les différentes isoformes de l'IL-17 est nécessaire à leur expression dans les lymphocytes Th17 et à l'ouverture de la chromatine (H3K4me3, 5hmC, p300) (280). SETDB1 étant un partenaire privilégié de TRIM28 et ayant été décrite comme capable de catalyser la marque H3K9me1, peu étudiée mais ayant été notamment associée aux promoteurs des gènes actifs (204), on peut donc également s'interroger sur le potentiel activateur de SETDB1. Il serait intéressant d'identifier le rôle de SETDB1 dans les autres lignages mais aussi de cartographier la distribution d'H3K9me1 dépendante de SETDB1.

En présence d'IL-12, les lymphocytes Th2 déficients pour SETDB1 sont capables d'allumer le programme d'expression génique Th1 mais aussi de réprimer le réseau de gènes Th2. Il serait intéressant de déterminer si la downmodulation du programme Th2 est un événement passif ou si des mécanismes actifs de répression, indépendants de SETDB1, sont impliqués. On peut imaginer que l'expression de Tbet inhibe le programme Th2 de par son interaction physique directe avec GATA3 (120). Il pourrait aussi réprimer l'expression du locus Th2 en se liant avec RUNX3 à l'élément répresseur HS IV (84). En effet, il a été montré que EZH2 et H3K27me3 s'accumulent sur HS IV dans les lymphocytes T de type Th1 (272).

## II. SETDB1 régule le programme Th1 via la déposition H3K9me3 sur un set restreint d'ERV

Nous avons montré que la délétion de SETDB1 s'associe à une perte drastique des pics d'H3K9me3 dans les lymphocytes T Th2, ce qui lève la répression du programme Th1. L'absence de répression provoque un défaut de stabilité des cellules qui, en présence de signaux pro-Th1, acquièrent le programme d'expression génique de type Th1. Nous avons donc investigué le rôle d'H3K9me3 dans la régulation du programme Th1.

### A. Propagation locale d'hétérochromatine

Dans les lymphocytes Th2 sauvages, nous avons observé qu'H3K9me3 est enrichie sur les ERV et n'est présente sur les enhancers Th1 que lorsque ceux-ci sont réprimés et flanqués d'un ERV. Le signal H3K9me3 étant centré sur les rétroéléments, nos résultats suggèrent donc que cette marque se propage des ERV jusqu'aux éléments cis-régulateurs adjacents. Nous avons donc émis l'hypothèse que la propagation d'H3K9me3 permet de réprimer les éléments régulateurs et les gènes situés à proximité. La propagation locale d'hétérochromatine comme mécanisme de répression de l'expression génique a déjà été proposée dans des mESC (222, 223). En tirant partie de deux lignées mESC n'ayant pas les mêmes IAP ou d'un modèle de mESC déficientes pour TRIM28, les auteurs de ces études ont montré qu'H3K9me3 se propage sur les gènes adjacents et joue un rôle causal dans leur répression.

### B. Evolution fonctionnelle des ERV en éléments cis-régulateurs

En plus d'un rôle potentiel dans la régulation des enhancers Th1 adjacents, les ERV pourraient eux-mêmes avoir évolué fonctionnellement en modules cis-régulateurs du programme Th1. Les éléments transposables (TE), de par leur structure, possèdent des sites de fixation à des facteurs de transcription. En effet, les TE utilisent la machinerie de transcription cellulaire pour induire l'expression de leurs gènes (290). La réplication et l'insertion des éléments transposables dans le génome constituent un risque mutagène pour l'organisme infecté. Les cellules hôtes ont donc mis en place des systèmes de répression. Les TE et l'organisme hôte ont alors co-évolué : les TE pour échapper à la répression et l'hôte pour limiter les risques. La coévolution des ERV et la recombinaison homologue qui peut s'effectuer entre deux LTR font que la plupart des ERV sont dégénérés et n'existent que sous la forme

d'un simple LTR dans le génome des mammifères. Néanmoins plusieurs études ont investigué les fonctions régulatrices de TE, qu'elles soient des fonctions d'enhancer, de promoteur ou de séquences isolatrices (« insulators »). Ces études se sont basées sur l'identification de signes caractéristiques d'éléments régulateurs comme la liaison de facteurs de transcription, un état de la chromatine spécifique du type cellulaire ou un impact de la dérégulation de TE spécifiques sur l'expression génique. Il a ainsi été montré que les sites de liaison à certains facteurs de transcription étaient fortement enrichis dans des familles de TE, essentiellement à LTR. En effet, ces derniers contiennent des séquences de liaison pour les facteurs de transcription qui sont sujettes à une forte pression de sélection (308, 309). De façon intéressante, Sundaram et collaborateurs ont montré que plus la famille de TE à LTR est liée par des facteurs de transcription, plus la chromatine est enrichie en marques permissives caractéristiques d'enhancers (H3K4me1 et H3K27ac) (310). De plus, il a été montré que la chromatine au niveau des TE adoptait une conformation permissive de façon spécifique du type cellulaire. Cette spécificité cellulaire est également retrouvée au niveau des processus biologiques associés aux gènes adjacents des TE, confortant ainsi l'hypothèse que les TE peuvent se comporter comme des enhancers (311). Dans la littérature, plusieurs exemples font cas de l'évolution fonctionnelle des ERV. Dans les cellules souches trophoblastiques, les ERVs de la famille RLTR13D5 peuvent fixer les facteurs de transcription EOMES, ELF5 et CDX2 et présentent les marques « d'enhancers » actifs H3K27ac et H3K4me1 (312). L'activation de ces ERV est, de plus, associée à l'expression de gènes trophoblastiques. Plus récemment, un rôle « enhancer » a été attribué à la famille d'ERV MER41 dans la régulation de AIM2, un composant inductible de l'inflammasome dans les macrophages humains (294). Ils ont en effet montré que les éléments MER41 sont ciblés par STAT1 grâce à une séquence consensus après stimulation des macrophages avec de l'IFN- $\gamma$ . Ces ERV qui fixent STAT1 sont enrichis à proximité de gènes stimulés par les interférons (ISG) et accumulent la marque permissive H3K27ac. La délétion de l'ERV MER41 associé à *Aim2* par un système CRISPR/Cas9 dans des cellules HELA abolit l'expression du gène en réponse à la stimulation cytokinique. Inversement, l'utilisation d'un système rapporteur luciférase a permis de montrer que cet élément était suffisant pour induire l'expression de la luciférase en réponse à une stimulation des cellules HELA par l'IFN- $\gamma$ . Si le groupe du Dr. Feschotte a ainsi mis en évidence le lien fonctionnel entre un ERV, ses caractéristiques d'enhancer et l'expression du gène cible, ils ont également identifié d'autres ERV de la famille MER41 qui ont évolué fonctionnellement en enhancers de gènes impliqués dans la réponse immunitaire. Ils ont donc révélé un réseau d'ERV aux fonctions d'enhancer permettant l'expression inductible de gènes en réponse à l'IFN- $\gamma$ . *In fine* on peut concevoir l'exaptation des ERVs comme un mécanisme des organismes hôtes pour détourner les ERV de leur potentiel mutagène à des fins bénéfiques pour les cellules ; et pour

les ERV, comme un mécanisme de détournement de la machinerie cellulaire de l'hôte afin d'assurer leur expression à défaut d'assurer leur réplication.

De façon cohérente, avec l'hypothèse que les ERV ont évolué fonctionnellement en modules *cis*-régulateurs du programme Th1, nous avons tout d'abord montré que les ERV sont marqués par H3K9me3 de façon spécifique du type cellulaire et sont associés à des processus biologiques impliqués dans la réponse immunitaire. De plus nous avons montré que les ERV sont enrichis en sites de liaison pour STAT1, STAT4, TBET, IRF1 et FOXO3 qui sont des facteurs de transcription impliqués dans l'induction et la stabilité du programme Th1. Le logiciel « Ingenuity Pathways Analysis » (IPA) a également identifié ces facteurs de transcription comme les plus à même d'expliquer les différences entre le transcriptome des lymphocytes T de type Th2 cultivés en condition pro-Th1 sauvage et celui des cellules déficientes pour SETDB1. Si nous n'avons pu effectuer l'analyse à l'échelle du génome de la distribution de STAT1 et de STAT4, nos analyses *in silico* ont révélé que les ERV marqués par H3K9me3 dans les lymphocytes Th2 intersectent des sites de fixation de STAT1 et de STAT4 identifiés dans des lymphocytes Th1 (62). De plus les ERV marqués par H3K9me3 dans les lymphocytes Th2 sont enrichis à proximité de gènes stimulés par les interférons (ISG) et chevauchent des « enhancers » actifs identifiés dans des lymphocytes Th1. Ceci suggère que lorsque les lymphocytes Th2 déficients pour SETDB1 sont soumis à des signaux pro-Th1, la signalisation du récepteur à l'IL-12 et du récepteur à l'IFN- $\gamma$  va induire l'activation de STAT4 et de STAT1 qui vont alors se fixer sur leurs éléments de réponse, notamment ceux compris dans les ERV. Ces ERV étant partiellement confondus avec des « enhancers » du programme Th1, ceci permet d'induire le programme d'expression génique Th1. De plus le logiciel IPA a identifié un réseau de gènes qui pourrait être contrôlé par des ERV ayant évolué fonctionnellement en « enhancers » du programme Th1. Cependant, nous n'avons pu établir un lien direct de cause à effet entre la perte de répression des ERV et l'expression du programme Th1.

### C. ERV module *cis*-régulateur ou DAMP

Si certains ERV ont évolué fonctionnellement en modules *cis*-régulateurs de l'expression génique, il est intéressant de noter qu'ils peuvent également servir de DAMP (Danger-Associated Molecule Pattern) à l'organisme lorsqu'ils sont exprimés. En effet, l'expression des ERV peut aboutir à la présence de divers acides nucléiques dérivés de rétrovirus qui vont être reconnus par des PRR (Pattern Recognition Receptors) variés. Dans les endosomes les TLR7 et 8 reconnaissent les ARN simple brin et le TLR3 reconnaît les ARN doubles brin. Les hélicases cytosoliques MDA-5 et RIG-1 sont, quant à elles, spécialisées dans la reconnaissance d'ARN double brin long et court respectivement. Le récepteur IFI16 (IFN- $\gamma$  inducible protein 16)

présent dans le cytoplasme et le nucléoplasme reconnaît l'ADN simple brin. Le récepteur cytosolique cGAS (cyclic GMP-AMP synthase), quant à lui, reconnaît l'ADN double brin. De plus, la reconnaissance d'ADN double brin dans le cytoplasme par AIM2 induit la formation d'inflammasome ; l'activation de la caspase-1 entraîne alors la production d'IL-1 $\beta$  et d'IL-18, des cytokines pro-inflammatoires, mais également la mort programmée de la cellule. L'activation des autres PRR résulte majoritairement en une réponse IFN de type I qui active le système immunitaire et induit l'expression des ISG.

La délétion de voies répressives de la chromatine peut induire l'expression d'ERV. Cela a été observé dans les lymphocytes B (225, 226) mais c'est également le cas dans des cellules cancéreuses traitées avec du 5-Aza-2-déoxycytidine pour inhiber la méthylation de l'ADN (292) ou déficientes pour SETDB1 (313). En revanche dans les lymphocytes Th2 déficients pour SETDB1, les ERV ayant perdu H3K9me3 ne sont pas transcrits ou ne le sont que faiblement. On peut imaginer que les facteurs de transcription nécessaires à leur expression ne sont pas présents ou que les ERV sont trop dégénérés pour être exprimés. Néanmoins, une faible minorité d'ERV fait exception et est exprimée. Il serait intéressant de comparer l'âge des ERV qui sont fortement transcrits à celui de ceux qui restent réprimés. On peut en effet imaginer que plus l'ERV est ancien, plus il est dégénéré ou plus les cellules auront mis en place des mécanismes de répression qui pourraient compenser l'absence de SETDB1. Malgré l'expression de certains ERV, nous n'avons pas non plus observé d'activation de voies MDA5/RIG-1 ni de surexpression des IFN de type I ou d'ISG. Ceci suggère que dans les lymphocytes Th2, H3K9me3 verrouille des ERV ayant évolué fonctionnellement en modules *cis*-régulateurs plutôt que des éléments transposables jouant le rôle de DAMPs ou qu'un autre mécanisme compensateur soit suffisant pour les réprimer.



### III. Ciblage de SETDB1 sur les ERV

Si SETDB1 est exprimée de façon ubiquitaire, elle dépose pourtant H3K9me3 sur un set restreint d'ERV qui est spécifique du type cellulaire ou du stade de différenciation. En effet, nous avons montré que les ERV marqués par H3K9me3 dans le tissu adipeux blanc ne sont pas les mêmes que ceux identifiés dans les lymphocytes Th2. On peut donc s'interroger sur les acteurs moléculaires permettant de cibler les ERV de façon si caractéristique du type cellulaire. Il a été montré que le complexe SETDB1 peut être recruté au niveau de l'euchromatine grâce à TRIM28 qui interagit avec des protéines à doigt de zinc et à domaines KRAB (KZFP) (211). Les KZFP possèdent une succession de doigts de zinc leur conférant une capacité de liaison à des séquences spécifiques de l'ADN. A l'inverse de TRIM28 qui est exprimé de façon ubiquitaire, l'expression des diverses KZFP est dépendante du type cellulaire et du stade de différenciation (295). La majorité de ces protéines se lie aux éléments transposables, en particulier les ERV. Elles ont co-évolué avec eux dans une dynamique similaire à celle « d'une course à l'armement », les KZFP afin d'endiguer l'invasion de ces parasites génomiques et les TE afin d'échapper à la répression des KZFP. En effet la plupart des KZFP sont apparues en même temps que les éléments transposables (ERV et LINE) qu'elles ciblent. Cependant, certaines semblent avoir été « détournées » de leur cible originelle pour réprimer de nouvelles vagues d'invasions rétrovirales (295). La pression de sélection qui s'opère au cours de l'évolution aurait permis de conserver les TE pouvant être reconnus par des KZFP et inversement les KZFP pouvant se lier à des TE. De façon cohérente avec le rôle émergent des ERV comme modules *cis*-régulateurs, des études de réseaux basées sur les sites de liaisons des KZFP et ceux d'acteurs moléculaires variés allant de facteur de transcription à des « chromatin writers » ont été réalisés. Elles ont révélé que la liaison des KZFP sur les TE coïncide avec celle de TRIM28, SETDB1 et HP1. Ceci suggère que les KZFP permettent de recruter des complexes répresseurs afin de former localement de l'hétérochromatine et verrouiller les TE. Les auteurs ont également identifié des KZFP dont les sites de liaison sur l'ADN correspondent à des sites de fixation à des facteurs de transcription. C'est par exemple le cas de la ZNF684 dont la liaison aux éléments MER41B, identifiés comme régulant les gènes inducibles par l'IFN- $\gamma$ , coïncide avec celle du facteur de transcription STAT1. Ceci suggère que les KZFP régulent des TE, notamment ceux pouvant fonctionner comme des modules *cis*-régulateurs (294). De plus il a été montré que la délétion de certaines KZFP est associée à la perte d'accumulation d'H3K9me3 sur les TE ciblés et à l'augmentation de marques d'enhancers actifs (H3K4me1 et H3K27ac). C'est par exemple le cas de ZFP809 qui est requise pour recruter KAP1 et SETDB1 sur les ERV possédant un site de liaison complémentaire à l'ARN de transfert de la proline (PBS-pro) et pour établir localement de l'hétérochromatine dans les fibroblastes embryonnaires

murins (314). Les KZFP apparaissent donc comme des candidates cohérentes pour déterminer la spécificité cellulaire de ciblage de SETDB1 sur les ERV.

A l'heure actuelle, le rôle des KZFP dans la biologie des lymphocytes T n'a pas été investigué. Néanmoins, l'analyse des transcriptomes de lymphocytes T naïfs, Th2 et Th2 soumis à des signaux polarisants Th1 pendant deux jours, que nous avons effectuée au laboratoire a permis d'identifier les sets de KZFP exprimées de façon différentielle selon le stade de différenciation des cellules et les stimuli auxquels elles ont été soumises. Plus de 15% des KZFP sont exprimées de façon différentielle, réprimées pour la majorité, au cours de la différenciation des cellules naïves en cellules Th2. La culture pendant deux jours des lymphocytes Th2 en présence d'IL-12 modifie de façon significative l'expression de 28 KZFP, pour moitié en la réprimant. Parmi ces KZFP exprimées de façon différentielle, nous avons identifié ZNF931 dont l'expression est réprimée dans les lymphocytes T Th2 soumis à des signaux polarisant Th1. De façon remarquable ZNF931, la protéine murine, est un orthologue de ZNF684 qui est la KZFP humaine pouvant se fixer sur les éléments MER41B, ces derniers ayant été identifiés comme régulant l'expression des gènes inductibles par l'IFN- $\gamma$  dans les monocytes (294, 295). Ceci laisse supposer que des KZFP orthologues peuvent avoir les mêmes fonctions dans des organismes distincts qui ne possèdent pas les mêmes TE.

## IV. Coopération des mécanismes épigénétiques dans la répression des ERV

Que les ERV aient évolué fonctionnellement en modules *cis*-régulateurs de l'expression génique, soient des DAMP ou plus simplement des « parasites » génomiques récemment intégrés, leur statut transcriptionnel doit être finement régulé et différents mécanismes épigénétiques coopèrent pour les réprimer. Si la voie PRC2/H3K27me3 n'a été que peu impliquée, les rôles de la méthylation de l'ADN et de la triméthylation d'H3K9 ont été décrits à de multiples reprises. Néanmoins, le champ d'action des différents acteurs moléculaires de ces deux voies ainsi que la coordination de ces deux mécanismes restent flous.

Nous avons mis en évidence dans les lymphocytes T Th2 que SETDB1 joue un rôle crucial, via la déposition d'H3K9me3, dans la répression d'ERV régulant le programme d'expression génique Th1. De façon remarquable, s'ils sont absents dans les lymphocytes T Th2 déficients pour SETDB1, les pics d'H3K9me3 ne disparaissent pas sur les ERV dans les cellules déficientes pour SUV39H1. Ceci révèle le rôle prépondérant de SETDB1 dans la répression de ces TE. Néanmoins, il a été suggéré dans les cellules embryonnaires murines que SETDB1 dépose H3K9me3 sur un set d'ERV à réprimer et que la marque est propagée par SUV39H1 sur les régions qui flanquent les ERV (224). On peut donc imaginer que si SETDB1 est nécessaire et semble suffisante pour réprimer les ERV, SUV39H1 vient consolider cette répression. En revanche dans les cellules souches embryonnaires murines SUV39H1 permet de réprimer les éléments transposables LINE, ces derniers perdent l'accumulation d'H3K9me3 en l'absence de la KMT et sont exprimés. Les différentes KMT permettant la déposition pourraient donc coopérer dans la répression des TE dans un même type cellulaire en ciblant des éléments différents.

Nous n'avons pas étudié le méthylome des lymphocytes T CD4 en absence ou en présence de SETDB1. Néanmoins, Thomas et collaborateurs ont identifié DNMT1 comme jouant un rôle crucial dans la répression de *Irfng* dans les lymphocytes T Th2. On peut donc s'interroger sur les cibles de DNMT1 dans ce contexte (269). De plus, seul 2% des ERV sont réprimés par H3K9me3 dans les cellules Th2 et en l'absence de SETDB1 certains ERV restent non exprimés. Si on peut imaginer que les facteurs de transcription nécessaires à leur expression ne sont pas présents ou que les ERV sont trop dégénérés pour être exprimés, on peut aussi émettre l'hypothèse que certains de ces ERV sont également réprimés par leur méthylation.

## V. Intérêts thérapeutiques de la modulation de l'axe SETDB1-H3K9me3-ERV

Les lymphocytes Th2, s'ils permettent de défendre l'organisme contre des parasites extracellulaires, sont également impliqués dans des pathologies inflammatoires chroniques tels la dermatite atopique, la colite à éosinophiles ou encore l'asthme allergique. Ces pathologies reposent sur un emballement de la réponse Th2. Par exemple, l'asthme allergique résulte de la stimulation répétée des lymphocytes Th2 en présence d'allergènes dans les bronches ce qui induit une forte production de cytokines (101). L'IL-4 et l'IL-13 induisent une bronchoconstriction des cellules musculaires et une production excessive de mucus par les cellules caliciformes alors que l'IL-5 favorise l'éosinophilie qui exacerbe les symptômes. Atténuer la réponse Th2 en la détournant vers une réponse Th1 peut être une piste thérapeutique envisageable pour lutter contre ces pathologies. On peut ainsi imaginer utiliser un inhibiteur de SETDB1 afin de permettre un biais Th1 dans la réponse CD4 et la diminution de la réponse inflammatoire de type Th2.

L'intérêt de moduler la réponse lymphocytaire apparaît également évident dans la lutte contre le cancer. En effet, l'accumulation de lymphocytes Th2 sur le site tumoral est majoritairement de mauvais pronostic vital, notamment dans les cas du cancer colorectal et du cancer gastrique (86, 315, 316). Le ratio des lymphocytes Th2 vs Th1 est d'ailleurs un marqueur prédictif du pronostic vital du patient (315), car à l'inverse une réponse lymphocytaire Th1 est presque systématiquement associée à un bon pronostic (**Figure 24**). En effet l'IFN- $\gamma$  produit peut agir directement sur les cellules tumorales pour inhiber leur prolifération et pour augmenter l'expression de molécules de CMH-I à leur surface ce qui augmente la probabilité qu'elles soient reconnues par des lymphocytes T CD8 cytotoxiques et lysées (88). L'IFN- $\gamma$  permet également d'amplifier les capacités cytotoxiques de ces derniers et des cellules NK. Il favorise aussi le recrutement de monocytes et leur activation en macrophages de type M1 produisant de grandes quantités de TNF- $\alpha$  et d'espèces réactives de l'oxygène. En agissant sur les cellules stromales, l'IFN- $\gamma$  altère également l'angiogenèse au niveau de la tumeur. Ainsi l'IFN- $\gamma$  permet d'établir un environnement inflammatoire propice à l'élimination des cellules cancéreuses. Inhiber la tri-méthylation d'H3K9 par SETDB1 lèverait un frein sur la production d'IFN- $\gamma$  par les LT de type Th1 et permettrait ainsi d'amplifier la réponse anti-tumorale.

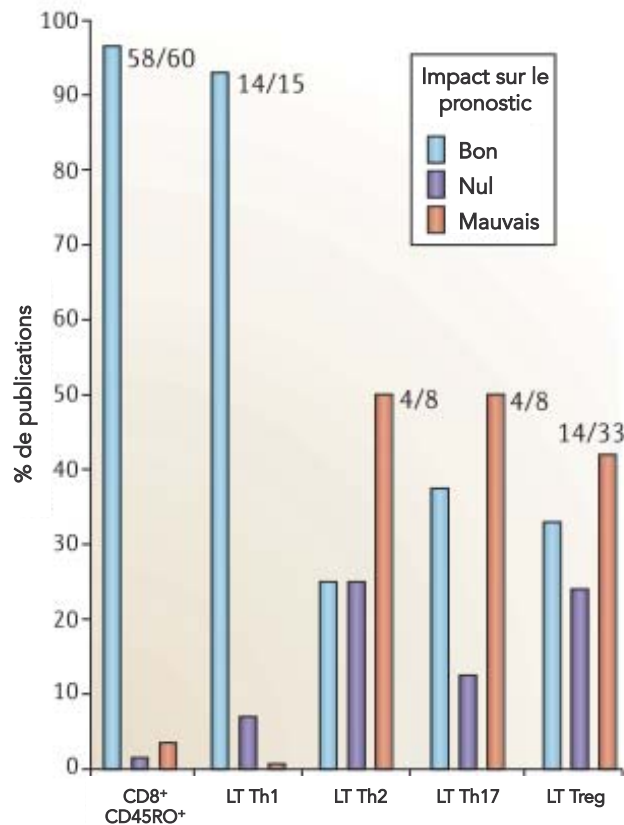


Figure 27 : Impact du type de lymphocytes T CD4 infiltrant les tumeurs sur le pronostic vital des patients

Les histogrammes représentent la proportion de publications ayant associé la présence de lymphocytes T infiltrant une tumeur (TIL) avec un pronostic vital du patient bon (bleu) ou mauvais (rouge) ainsi que la proportion de publications dans lesquelles la présence de TIL n'a pas d'impact sur le pronostic (violet). Si la présence de lymphocytes T CD8 et de lymphocytes T CD4 de type Th1 est associée à un pronostic favorable, l'impact des lymphocytes T de type Th17 et Treg est plus ambiguë et la présence de lymphocytes de type Th2 est majoritairement associée à un pronostic défavorable pour la survie du patient. Adapté de (86).

L'utilisation d'inhibiteur de SETDB1 pourrait également permettre de détourner le phénotype des lymphocytes Th2 vers le lignage Th1. Nous avons en effet mis en évidence que SETDB1, via la déposition d'H3K9me3, verrouille des ERV qui se comportent comme des éléments *cis*-régulateurs ou régulent des enhanceurs du programme Th1. Cependant le programme Th1 n'est exprimé dans les lymphocytes Th2 déficients pour SETDB1 que lorsqu'ils sont soumis à des signaux pro-Th1. On peut donc imaginer qu'il serait nécessaire de combiner l'inhibition de SETDB1 avec des signaux inflammatoires de type 1.

Les cellules cancéreuses présentent souvent des dérégulations d'enzymes impliquées dans la régulation des marques épigénétiques. Ces enzymes, qu'elles soient impliquées dans la méthylation de l'ADN, la déposition d'H3K27me3 ou d'H3K9me3, peuvent donc constituer des cibles en immunothérapie (317). En ce qui concerne SETDB1, elle est amplifiée dans des mélanomes et accélère leur survenue (318). Cuellar et collaborateurs ont également montré que SETDB1 joue un rôle majeur dans la leucémie myéloïde aigue (AML) (313). En effet en effectuant un criblage, grâce à la technologie d'édition du génome CRISPR/Cas9, de molécules éventuellement impliquées dans la survie des cellules AML, les auteurs ont ainsi identifié SETDB1 comme un acteur majeur. Dans les cellules AML, la délétion de SETDB1 induit une diminution partielle de la méthylation d'H3K9. Néanmoins la marque est perdue sur des gènes codant pour des KZFP et des éléments répétés. L'absence de répression de ces derniers résulte en une production d'ARN double brin qui active les voies RIG-I et MDA5. La signalisation de ces PRR permet d'induire une réponse antivirale reposant sur l'expression de gènes stimulés par les interférons (ISG) dont les *IFIT*. Le niveau d'H3K9me3 n'est pourtant pas diminué sur ces gènes, suggérant que leur expression n'est pas directement due à la délétion de SETDB1 mais pourrait être dépendante de la production d'IFN de type I en aval des voies RIG-1 et MDA5.

L'inhibition de SETDB1 dans un contexte tumoral pourrait agir comme une épée à double-tranchant. Cette stratégie permettrait d'une part, de façon intrinsèque aux cellules T, d'amplifier la réponse Th1 ou de détourner la programmation Th2 vers un phénotype Th1, et d'autre part, de déréguler la survie et/ou l'expression d'éléments répétés dans les cellules cancéreuses, boostant ainsi la réponse anti-tumorale en activant la sécrétion d'IFN de type I.

Chiappinelli et collaborateurs ont montré que traiter des lignées cellulaires issues de cancers ovariens avec le 5-Aza-2-deoxycytidine (Aza), un analogue de base ne permettant pas la méthylation de l'ADN, permettait d'induire l'expression et la sécrétion d'IFN- $\beta$  (292). Le traitement de lignées cancéreuses avec le surnageant des cellules traitées à l'Aza est suffisant pour induire une réponse antivirale et l'expression d'ISG. Une production d'IFN de type I par les cellules cancéreuses réduirait le besoin potentiel, évoqué précédemment, de combiner des signaux pro-Th1 avec des inhibiteurs de SETDB1 pour détourner la polarisation des lymphocytes T Th2 vers le phénotype Th1. L'augmentation des défenses virales suite au traitement des tumeurs primaires corrèle positivement avec la dérégulation des ERV mais surtout avec la sensibilité tumorale au traitement, faisant des traitements à base d'inhibiteurs de mécanismes répresseurs des ERV, des candidats intéressants en thérapie. Néanmoins, ces molécules étant souvent ubiquitaires, leur manipulation est risquée car elles sont susceptibles de générer de nombreux effets secondaires sur les tissus sains.

# REFERENCES

1. E. V. Rothenberg, J. E. Moore, M. A. Yui, Launching the T-cell-lineage developmental programme. *Nat. Rev. Immunol.* **8**, 9–21 (2008).
2. M. A. Yui, E. V. Rothenberg, Developmental gene networks: A triathlon on the course to T cell identity. *Nat. Rev. Immunol.* **14**, 529–545 (2014).
3. K. Lucas, D. Vremec, L. Wu, K. Shortman, A linkage between dendritic cell and T-cell development in the mouse thymus: The capacity of sequential T-cell precursors to form dendritic cells in culture. *Dev. Comp. Immunol.* **22**, 339–349 (1998).
4. K. Masuda *et al.*, T Cell Lineage Determination Precedes the Initiation of TCR Gene Rearrangement. *J. Immunol.* **179**, 3699–3706 (2007).
5. J. L. Johnson *et al.*, Lineage-Determining Transcription Factor TCF-1 Initiates the Epigenetic Identity of T Cells Article Lineage-Determining Transcription Factor TCF-1 Initiates the Epigenetic Identity of T Cells. *Immunity*. **48**, 243–257.e10 (2018).
6. C. C. Tydell *et al.*, Molecular Dissection of Prethymic Progenitor Entry into the T Lymphocyte Developmental Pathway. *J. Immunol.* **179**, 421–438 (2007).
7. L. Li, M. Leid, E. V. Rothenberg, An Early T Cell Lineage Commitment Checkpoint Dependent on the Transcription Factor Bcl11b. *Science (80-. )*. **329**, 89–93 (2010).
8. V. V. Kapitonov, J. Jurka, RAG1 core and V(D)J recombination signal sequences were derived from Transib transposons. *PLoS Biol.* **3**, 0998–1011 (2005).
9. S. D. Fugmann, The origins of the Rag genes- From transposition to V(D)J recombination. *Semin. Immunol.* **22**, 10–16 (2010).
10. P. Mombaerts *et al.*, RAG-1-deficient mice have no mature B and T lymphocytes. *Cell*. **68**, 869–877 (1992).
11. Y. Shinkai *et al.*, RAG-2-deficient mice lack mature lymphocytes owing to inability to initiate V(D)J rearrangement. *Cell*. **68**, 855–867 (1992).
12. C. Saint-Ruf *et al.*, Analysis and expression of a cloned pre-T cell receptor gene. *Science*. **266**, 1208–1212 (1994).
13. M. Groettrup *et al.*, A novel disulfide-linked heterodimer on pre-T cells consists of the T cell receptor  $\beta$  chain and a 33 kd glycoprotein. *Cell*. **75**, 283–294 (1993).
14. L. Wu, C.-L. Li, K. Shortman, Thymic dendritic cell precursors: relationship to the T lymphocyte lineage and phenotype of the dendritic cell progeny. *J. Exp. Med.* **184** (1996).
15. M. Merckenschlager *et al.*, How Many Thymocytes Audition for Selection? *J. Exp. Med.* **186**, 1149–1158 (1997).
16. T. M. Laufer, J. DeKoning, J. S. Markowitz, D. Lo, L. H. Glimcher, Unopposed positive selection and autoreactivity in mice expressing class II MHC only on thymic cortex. *Nature*. **383** (1996), pp. 81–85.
17. T. K. Starr, S. C. Jameson, K. A. Hogquist, POSITIVE AND NEGATIVE SELECTION OF TCELLS. *Annu. Rev. Immunol.* **21**, 139–176 (2003).
18. R. Bosselut, CD4/CD8-lineage differentiation in the thymus: from nuclear effectors to membrane signals. *Nat. Rev. Immunol.* **4**, 529–540 (2004).
19. R. Förster, A. C. Davalos-Misilitz, A. Rot, CCR7 and its ligands: Balancing immunity and tolerance. *Nat. Rev. Immunol.* **8**, 362–371 (2008).
20. M. S. Anderson *et al.*, Projection of an immunological self shadow within the thymus by the aire protein. *Science (80-. )*. **298**, 1395–1401 (2002).
21. S. Z. Josefowicz, L.-F. Lu, A. Y. Rudensky, Regulatory T Cells: Mechanisms of Differentiation and Function. *Annu. Rev. Immunol.* **30**, 531–564 (2012).
22. H. Nagaoka, G. Gonzalez-Aseguinolaza, M. Tsuji, M. C. Nussenzweig, Immunization and infection change the number of recombination activating gene (RAG)-expressing B cells in the periphery by altering immature lymphocyte production. *J. Exp. Med.* **191**, 2113–20 (2000).
23. T. E. Boursalian, J. Golob, D. M. Soper, C. J. Cooper, P. J. Fink, Continued maturation of thymic emigrants in the periphery. *Nat. Immunol.* **5**, 418–425 (2004).
24. T. M. McCaughtry, M. S. Wilken, K. A. Hogquist, Thymic emigration revisited. *J. Exp. Med.* **204**, 2513–2520 (2007).
25. D. W. Hendricks, P. J. Fink, *Blood*, in press, doi:10.1182/blood-2010-07-299263.
26. P. Friedl, A. T. Den Boer, M. Gunzer, Tuning immune responses: Diversity and adaptation of the immunological synapse. *Nat. Rev. Immunol.* **5**, 532–545 (2005).



27. T. Brocker, Survival of mature CD4 T lymphocytes is dependent on major histocompatibility complex class II-expressing dendritic cells. *J. Exp. Med.* **186**, 1223–32 (1997).
28. L. Vivien, C. Benoist, D. Mathis, T lymphocytes need IL-7 but not IL-4 or IL-6 to survive in vivo. *Int. Immunol.* **13**, 763–768 (2001).
29. L. Onder et al., IL-7 – producing stromal cells are critical for lymph node remodeling. *Blood.* **120**, 4675–4683 (2012).
30. C. D. Surh, J. Sprent, Homeostasis of Naive and Memory T Cells. *Immunity.* **29**, 848–862 (2008).
31. S. Wojciechowski et al., Bim/Bcl-2 balance is critical for maintaining naive and memory T cell homeostasis. *J. Exp. Med.* **204**, 1665–1675 (2007).
32. N. J. Maciver, R. D. Michalek, J. C. Rathmell, Metabolic Regulation of T Lymphocytes. *Annu. Rev. Immunol.* **31**, 259–283 (2013).
33. J. A. Wofford, H. L. Wieman, S. R. Jacobs, Y. Zhao, J. C. Rathmell, IL-7 promotes Glut1 trafficking and glucose uptake via STAT5-mediated activation of Akt to support T cell survival. *Blood.* **111**, 2101–2112 (2008).
34. U. H. Von Andrian, T. R. Mempel, Homing and cellular traffic in lymph nodes. *Nat. Rev. Immunol.* **3**, 867–878 (2003).
35. K. Takada, S. C. Jameson, Naive T cell homeostasis: From awareness of space to a sense of place. *Nat. Rev. Immunol.* **9**, 823–832 (2009).
36. D. L. Mueller, Mechanisms maintaining peripheral tolerance. *Nat. Immunol.* **11**, 21–27 (2010).
37. C. G. Fathman, N. B. Lineberry, Molecular mechanisms of CD4+ T-cell anergy. *Nat. Rev. Immunol.* **7**, 599–609 (2007).
38. D. Hawiger et al., Dendritic Cells Induce Peripheral T Cell Unresponsiveness under Steady State Conditions in Vivo. *J. Exp. Med.* **194**, 769–780 (2001).
39. O. P. Joffre, D. Sancho, S. Zelenay, A. M. Keller, C. Reis E Sousa, Efficient and versatile manipulation of the peripheral CD4+ T-cell compartment by antigen targeting to DNGR-1/CLEC9A. *Eur. J. Immunol.* **40**, 1255–1265 (2010).
40. S. Xiao et al., Retinoic Acid Increases Foxp3+ Regulatory T Cells and Inhibits Development of Th17 Cells by Enhancing TGF $\beta$  Driven Smad3 Signaling and Inhibiting IL-6 and IL-23 Receptor Expression. *J. Immunol.* **181**, 2277–2284 (2008).
41. D. Mucida et al., Reciprocal TH17 and regulatory T cell differentiation mediated by retinoic acid. *Science (80-. ).* **317**, 256–260 (2007).
42. M. E. Brunkow et al., Disruption of a new forkhead/winged-helix protein, scurfy, results in the fatal lymphoproliferative disorder of the scurfy mouse. *Nat. Genet.* **27**, 68–73 (2001).
43. C. L. Bennett et al., The immune dysregulation, polyendocrinopathy, enteropathy, X-linked syndrome (IPEX) is caused by mutations of FOXP3. *Nat. Genet.* **27**, 20–21 (2001).
44. D. A. A. Vignali, L. W. Collison, C. J. Workman, How regulatory T cells work. *Nat. Rev. Immunol.* **8**, 523–532 (2008).
45. F. Fallarino et al., Modulation of tryptophan catabolism by regulatory T cells. *Nat. Immunol.* **4**, 1206–1212 (2003).
46. C. A. Janeway, R. Medzhitov, Innate Immune Recognition. *Annu. Rev. Immunol.* **20**, 197–216 (2002).
47. X. Cao, Self-regulation and cross-regulation of pattern-recognition receptor signalling in health and disease. *Nat. Rev. Immunol.* **16**, 35–50 (2016).
48. P. Matzinger, An innate sense of danger. *Semin. Immunol.* **10**, 399–415 (1998).
49. S. Y. Seong, P. Matzinger, Hydrophobicity: An ancient damage-associated molecular pattern that initiates innate immune responses. *Nat. Rev. Immunol.* **4**, 469–478 (2004).
50. D. N. J. Hart, Dendritic cells: unique leukocytes populations which control the primary immune response. *Blood.* **90**, 3245–3287 (1997).
51. T. R. Mempel, S. E. Henrickson, U. H. Von Andrian, T-cell priming by dendritic cells in lymph nodes occurs in three distinct phases. *Nature.* **427**, 154–159 (2004).
52. L. Chen, D. B. Flies, Molecular mechanisms of T cell co-stimulation and co-inhibition. *Nat. Rev. Immunol.* **13**, 227–242 (2013).
53. R. A. Maldonado et al., Control of T helper cell differentiation through cytokine receptor inclusion in the immunological synapse. *J. Exp. Med.* **206**, 877–892 (2009).
54. G. Gaud, R. Lesourne, P. E. Love, Regulatory mechanisms in T cell receptor signalling. *Nat. Rev. Immunol.*, 1–13 (2018).
55. L. H. Boise et al., CD28 costimulation can promote T cell survival by enhancing the expression of Bcl-xL. *Immunity.* **3**, 87–98 (1995).



56. C. Cerdan *et al.*, Prolonged IL-2 receptor alpha/CD25 expression after T cell activation via the adhesion molecules CD2 and CD28. Demonstration of combined transcriptional and post-transcriptional regulation. *J. Immunol.* **149**, 2255–2261 (1992).
57. C. B. Thompson *et al.*, CD28 activation pathway regulates the production of multiple T-cell-derived lymphokines/cytokines. *Proc. Natl. Acad. Sci.* **86**, 1333–1337 (1989).
58. L. J. Appleman, A. Berezovskaya, I. Grass, V. A. Boussiotis, CD28 costimulation mediates T cell expansion via IL-2-independent and IL-2-dependent regulation of cell cycle progression. *J. Immunol.* **164**, 144–151 (2000).
59. K. A. Frauwirth *et al.*, The CD28 signaling pathway regulates glucose metabolism. *Immunity.* **16**, 769–777 (2002).
60. J. Hendriks *et al.*, CD27 is required for generation and long-term maintenance of T cell immunity. *Nat. Immunol.* **1**, 433–440 (2000).
61. J. Hendriks, Y. Xiao, J. Borst, CD27 Promotes Survival of Activated T Cells and Complements CD28 in Generation and Establishment of the Effector T Cell Pool. *J. Exp. Med.* **198**, 1369–1380 (2003).
62. G. Vahedi *et al.*, STATs shape the active enhancer landscape of T cell populations. *Cell.* **151**, 981–993 (2012).
63. K. Hirahara *et al.*, Asymmetric Action of STAT Transcription Factors Drives Transcriptional Outputs and Cytokine Specificity. *Immunity.* **42**, 877–889 (2015).
64. D. Amsen, A. Antov, R. A. Flavell, The different faces of Notch in T-helper-cell differentiation. *Nat. Rev. Immunol.* **9**, 116–124 (2009).
65. N. van Panhuys, F. Klauschen, R. N. Germain, T-Cell-Receptor-Dependent Signal Intensity Dominantly Controls CD4+ T Cell Polarization In Vivo. *Immunity.* **41**, 63–74 (2014).
66. H. Soares *et al.*, A subset of dendritic cells induces CD4<sup>+</sup> T cells to produce IFN- $\gamma$  by an IL-12-independent but CD70-dependent mechanism in vivo. *J. Exp. Med.* **204**, 1095–1106 (2007).
67. J. M. Coquet *et al.*, The CD27 and CD70 Costimulatory Pathway Inhibits Effector Function of T Helper 17 Cells and Attenuates Associated Autoimmunity. *Immunity.* **38**, 53–65 (2013).
68. J. J. Bird *et al.*, Helper T cell differentiation is controlled by the cell cycle. *Immunity.* **9**, 229–237 (1998).
69. J. L. Grogan, M. Mohrs, B. Harmon, D. A. Lacy, J. W. Sedat, Early transcription and silencing of cytokine genes underlie polarization of T helper cell subsets. *Immunity* (2001).
70. T. R. Mosmann, H. Cherwinski, M. W. Bond, M. A. Giedlin, R. L. Coffman, Two types of murine helper T cell clone. I. Definition according to profiles of lymphokine activities and secreted proteins. *J. Immunol.* **136**, 2348–57 (1986).
71. F. Koch *et al.*, High level IL-12 production by murine dendritic cells: upregulation via MHC class II and CD40 molecules and downregulation by IL-4 and IL-10. *J. Exp. Med.* **184**, 741–746 (1996).
72. B. M. Celia *et al.*, Ligation of CD40 on Dendritic Cells Triggers Production of High Levels of Interleukin-12 and Enhances T Cell Stimulatory Capacity: T-T Help via APC Activation. *J. Exp. Med.* **184**, 746–752 (1996).
73. C. Y. Wu, M. Gadina, K. Wang, J. O'Shea, R. a Seder, Cytokine regulation of IL-12 receptor beta2 expression: differential effects on human T and NK cells. *Eur. J. Immunol.* **30**, 1364–74 (2000).
74. W. Liao, J. X. Lin, L. Wang, P. Li, W. J. Leonard, Modulation of cytokine receptors by IL-2 broadly regulates differentiation into helper T cell lineages. *Nat. Immunol.* **12**, 551–559 (2011).
75. L. L. Carter, K. M. Murphy, Lineage-specific requirement for signal transducer and activator of transcription (Stat)4 in interferon gamma production from CD4(+) versus CD8(+) T cells. *J. Exp. Med.* **189**, 1355–60 (1999).
76. M. H. Kaplan, Y. L. Sun, T. Hoey, M. J. Grusby, Impaired IL-12 responses and enhanced development of Th2 cells in Stat4-deficient mice. *Nature.* **382** (1996), pp. 174–177.
77. W. E. Thierfelder *et al.*, Requirement for Stat4 in interleukin-12-mediated responses of natural killer and T cells. *Nature.* **382** (1996), pp. 171–174.
78. J. Zhu *et al.*, The Transcription Factor T-bet Is Induced by Multiple Pathways and Prevents an Endogenous Th2 Cell Program during Th1 Cell Responses. *Immunity.* **37**, 660–673 (2012).
79. S. J. Szabo *et al.*, A novel transcription factor, T-bet, directs Th1 lineage commitment. *Cell.* **100**, 655–669 (2000).
80. S. J. Szabo *et al.*, Distinct Effects of T-bet in T

- H 1 Lineage Commitment and IFN- $\gamma$  Production in CD4 and CD8 T Cells. *Science* (80-. ). **295**, 338–343 (2002).
81. A. Martín-Fontecha et al., Induced recruitment of NK cells to lymph nodes provides IFN- $\gamma$  for TH1 priming. *Nat. Immunol.* **5**, 1260–1265 (2004).
82. M. DuPage, J. A. Bluestone, Harnessing the plasticity of CD4+ T cells to treat immune-mediated disease. *Nat. Rev. Immunol.* **16**, 149–163 (2016).
83. A. C. Mullen et al., Hlx is induced by and genetically interacts with T-bet to promote heritable TH1 gene induction. *Nat. Immunol.* **3**, 652–658 (2002).
84. I. M. Djuretic et al., Transcription factors T-bet and Runx3 cooperate to activate Ifng and silence Il4 in T helper type 1 cells. *Nat. Immunol.* **8**, 145–53 (2007).
85. R. Bonecchi et al., Differential Expression of Chemokine Receptors and Chemotactic Responsiveness of Type 1 T Helper Cells (Th1s) and Th2s. *J. Exp. Med.* **187**, 129–134 (1998).
86. W. H. Fridman, F. Pagès, C. Sautès-Fridman, J. Galon, The immune contexture in human tumours: Impact on clinical outcome. *Nat. Rev. Cancer.* **12**, 298–306 (2012).
87. C. F. Nathan, H. W. Murray, I. E. Wlebe, B. Y. Rubin, Identification of interferon-gamma as the lymphokine that activates human macrophage oxidative metabolism and antimicrobial activity. *J. exp.* **158**, 670–689 (1983).
88. L. B. Ivashkiv, IFN $\gamma$ : signalling, epigenetics and roles in immunity, metabolism, disease and cancer immunotherapy. *Nat. Rev. Immunol.*, 1 (2018).
89. P. Keskinen, T. Ronni, S. Matikainen, A. Lehtonen, I. Julkunen, Regulation of HLA class I and II expression by interferons and influenza A virus in human peripheral blood mononuclear cells. *Immunology.* **91**, 421–429 (1997).
90. G. Le Gros, S. Z. Ben-sasson, R. Seder, F. D. Finkelman, W. E. Paul, Generation of Interleukin 4 (IL-4)-producing Cells In Vivo and In Vitro: IL-2 and IL-4 Are Required For In Vitro Generation of IL-4-producing Cells. *J. Exp. Med.* **172**, 921–929 (1990).
91. S. L. Swain, A. D. Weinberg, M. English, G. Huston, G. Huston, IL-4 directs the development of Th2-like helper effectors . *J. Immunol.* **145**, 3796–3806 (1990).
92. B. Min et al., Basophils Produce IL-4 and Accumulate in Tissues after Infection with a Th2-inducing Parasite. *J. Exp. Med.* **200**, 507–517 (2004).
93. Y.-L. Ma et al., IL-4–Producing Dendritic Cells Induced during *Schistosoma japonica* Infection Promote Th2 Cells via IL-4–Dependent Pathway. *J. Immunol.* **195**, 3769–3780 (2015).
94. R. M. Anthony, L. I. Rutitzky, J. F. Urban, M. J. Stadecker, W. C. Gause, Protective immune mechanisms in helminth infection. *Nat. Rev. Immunol.* **7**, 975–987 (2007).
95. M. H. Kaplan, U. Schindler, S. T. Smiley, M. J. Grusby, Stat6 is required for mediating responses to IL-4 and for the development of Th2 cells. *Immunity.* **4**, 313–319 (1996).
96. H. Kurata, H. J. Lee, A. O’Garra, N. Arai, Ectopic expression of activated Stat6 induces the expression of Th2-specific cytokines and transcription factors in developing Th1 cells. *Immunity.* **11**, 677–688 (1999).
97. W. Zheng, R. A. Flavell, The transcription factor GATA-3 is necessary and sufficient for Th2 cytokine gene expression in CD4 T cells. *Cell.* **89**, 587–596 (1997).
98. B. I. Ho, D. Lo, L. H. Glimcher, c-maf Promotes T Helper Cell Type 2 (Th2) and Attenuates Th1 Differentiation by Both Interleukin 4–dependent and –independent Mechanisms. *J. Exp. Med.* **188**, 1859–1866 (1998).
99. J. E. Allen, R. M. Maizels, Diversity and dialogue in immunity to helminths. *Nat. Rev. Immunol.* **11**, 375–388 (2011).
100. S. S. Possa, E. A. Leick, C. M. Prado, M. A. Martins, I. F. L. C. Tibério, Eosinophilic inflammation in allergic asthma. *Front. Pharmacol.* **4**, 1–9 (2013).
101. B. N. Lambrecht, H. Hammad, The immunology of asthma. *Nat. Immunol.* **16**, 45–56 (2015).
102. B. L. Kelly, R. M. Locksley, Coordinate Regulation of the IL-4, IL-13, and IL-5 Cytokine Cluster in Th2 Clones Revealed by Allelic Expression Patterns. *J. Immunol.* **165**, 2982–2986 (2000).
103. L. Zhou et al., IL-6 programs TH-17 cell differentiation by promoting sequential engagement of the IL-21 and IL-23 pathways. *Nat. Immunol.* **8**, 967–974 (2007).
104. S. Aggarwal, N. Ghilardi, M. H. Xie, F. J. De Sauvage, A. L. Gurney, Interleukin-23 promotes a distinct CD4 T cell activation state characterized by the production of interleukin-17. *J. Biol. Chem.* **278**, 1910–1914 (2003).
105. L. E. Harrington et al., Interleukin 17–

- producing CD4+ effector T cells develop via a lineage distinct from the T helper type 1 and 2 lineages. *Nat. Immunol.* **6**, 1123–1132 (2005).
106. P. Muranski, N. P. Restifo, W. Dc, Essentials of Th17 cell commitment and plasticity Review Article Essentials of Th17 cell commitment and plasticity. *Blood*. **121**, 2402–2414 (2013).
  107. I. I. Ivanov et al., The orphan nuclear receptor ROR $\gamma$  directs the differentiation program of proinflammatory IL-17+ T helper cells. *Cell*. **126**, 1121–1133 (2006).
  108. S. C. Liang et al., Interleukin (IL)-22 and IL-17 are coexpressed by Th17 cells and cooperatively enhance expression of antimicrobial peptides. *J. Exp. Med.* **203**, 2271–2279 (2006).
  109. M. J. McGeachy et al., TGF- $\beta$  and IL-6 drive the production of IL-17 and IL-10 by T cells and restrain TH-17 cell-mediated pathology. *Nat. Immunol.* **8**, 1390–1397 (2007).
  110. L. A. Zúñiga, R. Jain, C. Haines, D. J. Cua, Th17 cell development: From the cradle to the grave. *Immunol. Rev.* **252**, 78–88 (2013).
  111. M. El-Behi et al., The encephalitogenicity of TH17 cells is dependent on IL-1- and IL-23-induced production of the cytokine GM-CSF. *Nat. Immunol.* **12**, 568–575 (2011).
  112. N. M. Haynes et al., Role of CXCR5 and CCR7 in Follicular Th Cell Positioning and Appearance of a Programmed Cell Death Gene-1<sup>High</sup> Germinal Center-Associated Subpopulation. *J. Immunol.* **179**, 5099–5108 (2007).
  113. H. Akiba et al., The Role of ICOS in the CXCR5+ Follicular B Helper T Cell Maintenance In Vivo. *J. Immunol.* **175**, 2340–2348 (2005).
  114. D. J. Campbell, C. H. Kim, E. C. Butcher, Separable effector T cell populations specialized for B cell help or tissue inflammation. *Nat Immunol.* **2**, 876–81. (2001).
  115. R. I. Nurieva et al., Generation of T Follicular Helper Cells Is Mediated by Interleukin-21 but Independent of T Helper 1, 2, or 17 Cell Lineages. *Immunity*. **29**, 138–149 (2008).
  116. R. I. Nurieva et al., Bcl6 Mediates the Development of T Follicular Helper Cells. *Science (80-. )*. **35**, 1001–1006 (2009).
  117. D. Yu et al., The Transcriptional Repressor Bcl-6 Directs T Follicular Helper Cell Lineage Commitment. *Immunity*. **31**, 457–468 (2009).
  118. T. R. Mosmann, H. Cherwinski, M. W. Bond, M. A. Giedlin, R. L. Coffman, Two types of murine helper T cell clone. I. Definition according to profiles of lymphokine activities and secreted proteins. *J. Immunol.* **136**, 2348–2357 (1986).
  119. B. J. Holaday et al., Reconstitution of Leishmania immunity in severe combined immunodeficient mice using Th1- and Th2-like cell lines. *J. Immunol.* **147**, 1653–1658 (1991).
  120. E. S. Hwang, S. J. Szabo, P. L. Schwartzberg, L. H. Glimcher, T helper cell fate specified by kinase-mediated interaction of T-bet with GATA-3. *Science (80-. )*. **307**, 430–433 (2005).
  121. T. Usui et al., T-bet regulates Th1 responses through essential effects on GATA-3 function rather than on IFNG gene acetylation and transcription. *J. Exp. Med.* **203**, 755–66 (2006).
  122. B. R. Manetti et al., Natural Killer Cell Stimulatory Factor (Interleukin 12 [IL-12]) Induces T Helper Type 1 (Th1)-specific Immune Responses and Inhibits the Development of IL-4-producing Th Cells. *J. Exp. Med.* **177**, 1199–1204 (1993).
  123. W. Ouyang et al., Inhibition of Th1 development mediated by GATA-3 through an IL-4- independent mechanism. *Immunity*. **9**, 745–755 (1998).
  124. T. Usui, R. Nishikomori, A. Kitani, W. Strober, GATA-3 suppresses Th1 development by downregulation of Stat4 and not through effects on IL-12R $\beta$ 2 chain or T-bet. *Immunity*. **18**, 415–428 (2003).
  125. C. A. Knosp, J. A. Johnston, Regulation of CD4+ T-cell polarization by suppressor of cytokine signalling proteins. *Immunology*. **135**, 101–111 (2011).
  126. A. Kelso, Th1 and Th2 subset: paradigms lost? *Immunol. Today*. **16**, 374–379 (1995).
  127. S. Sakaguchi, N. Sakaguchi, M. Asano, M. Itoh, M. Toda, Immunologic Self-Tolerance Maintained by Activated T Cells Expressing 11-2 Receptor  $\alpha$ -Chains (CD25). *J. Immunol.* **155**, 1151–1164 (1995).
  128. E. Schmitt, M. Klein, T. Bopp, Th9 cells, new players in adaptive immunity. *Trends Immunol.* **35**, 61–68 (2014).
  129. M. Veldhoen et al., Transforming growth factor- $\beta$  “reprograms” the differentiation of T helper 2 cells and promotes an interleukin 9-producing subset. *Nat. Immunol.* **9**, 1341–1346 (2008).
  130. N. Gagliani et al., Th17 cells transdifferentiate into regulatory T cells during resolution of inflammation. *Nature*. **523**, 221–5 (2015).

131. Y. K. Lee *et al.*, Late Developmental Plasticity in the T Helper 17 Lineage. *Immunity*. **30**, 92–107 (2009).
132. S. N. Harbour, C. L. Maynard, C. L. Zindl, T. R. Schoeb, C. T. Weaver, Th17 cells give rise to Th1 cells that are required for the pathogenesis of colitis. *Proc. Natl. Acad. Sci.* **112**, 7061–7066 (2015).
133. Y.-H. Wang *et al.*, A novel subset of CD4<sup>+</sup> T<sub>H</sub>2 memory/effector cells that produce inflammatory IL-17 cytokine and promote the exacerbation of chronic allergic asthma. *J. Exp. Med.* **207**, 2479–2491 (2010).
134. L. Cosmi *et al.*, Identification of a novel subset of human circulating memory CD4<sup>+</sup> T cells that produce both IL-17A and IL-4. *J. Allergy Clin. Immunol.* **125**, 222–230 (2010).
135. M. Raymond, V. Q. Van, K. Wakahara, M. Rubio, M. Sarfati, Lung dendritic cells induce TH17 cells that produce TH2 cytokines, express GATA-3, and promote airway inflammation. *J. Allergy Clin. Immunol.* **128**, 192–201 (2011).
136. K. T. Lu *et al.*, Functional and Epigenetic Studies Reveal Multistep Differentiation and Plasticity of In Vitro-Generated and In Vivo-Derived Follicular T Helper Cells. *Immunity*. **35**, 622–632 (2011).
137. R. L. Reinhardt, H. E. Liang, R. M. Locksley, Cytokine-secreting follicular T cells shape the antibody repertoire. *Nat. Immunol.* **10**, 385–393 (2009).
138. T. Feng, A. T. Cao, C. T. Weaver, C. O. Elson, Y. Cong, Interleukin-12 converts Foxp3<sup>+</sup> regulatory T cells to interferon- $\gamma$ -producing Foxp3<sup>+</sup> T cells that inhibit colitis. *Gastroenterology*. **140**, 2031–2043 (2011).
139. M. A. Koch *et al.*, The transcription factor T-bet controls regulatory T cell homeostasis and function during type 1 inflammation. *Nat. Immunol.* **10**, 595–602 (2009).
140. Y. Wang, M. A. Su, Y. Y. Wan, An Essential Role of the Transcription Factor GATA-3 for the Function of Regulatory T Cells. *Immunity*. **35**, 337–348 (2011).
141. Y. Chung *et al.*, Follicular regulatory T cells expressing Foxp3 and Bcl-6 suppress germinal center reactions. *Nat. Med.* **17**, 983–988 (2011).
142. J. Hua *et al.*, Pathological conversion of regulatory T cells is associated with loss of allotolerance. *Sci. Rep.* **8**, 1–9 (2018).
143. G. Oldenhove *et al.*, Decrease of Foxp3<sup>+</sup>Treg Cell Number and Acquisition of Effector Cell Phenotype during Lethal Infection. *Immunity*. **31**, 772–786 (2009).
144. A. N. Hegazy *et al.*, Interferons Direct Th2 Cell Reprogramming to Generate a Stable GATA-3<sup>+</sup>T-bet<sup>+</sup> Cell Subset with Combined Th2 and Th1 Cell Functions. *Immunity*. **32**, 116–128 (2010).
145. M. Peine *et al.*, Stable T-bet+GATA-3+ Th1/Th2 Hybrid Cells Arise In Vivo, Can Develop Directly from Naive Precursors, and Limit Immunopathologic Inflammation. *PLoS Biol.* **11** (2013), doi:10.1371/journal.pbio.1001633.
146. S. M. Coomes *et al.*, IFN $\gamma$  and IL-12 Restrict Th2 Responses during Helminth/Plasmodium Co-Infection and Promote IFN $\gamma$  from Th2 Cells. *PLoS Pathog.* **11**, 1–29 (2015).
147. A. P. Wolffe, H. Kurumizaka, The Nucleosome: A Powerful Regulator of Transcription. *Prog. Nucleic Acid Res. Mol. Biol.* **61**, 379–422 (1998).
148. G. Felsenfeld, M. Groudine, Controlling the double helix. *Nature*. **421**, 448–453 (2003).
149. P. Trojer, D. Reinberg, Facultative Heterochromatin: Is There a Distinctive Molecular Signature? *Mol. Cell.* **28**, 1–13 (2007).
150. S. I. S. Grewal, S. Jia, Heterochromatin revisited. *Nat. Rev. Genet.* **8**, 35–46 (2007).
151. R. C. Allshire, H. D. Madhani, Ten principles of heterochromatin formation and function. *Nat. Rev. Mol. Cell Biol.* **19**, 229–244 (2018).
152. T. Juven-Gershon, J. T. Kadonaga, Regulation of gene expression via the core promoter and the basal transcriptional machinery. *Dev. Biol.* **339**, 225–229 (2010).
153. C. G. Spilianakis, R. A. Flavell, Long-range intrachromosomal interactions in the T helper type 2 cytokine locus. *Nat. Immunol.* **5**, 1017–1027 (2004).
154. D. Hnisz *et al.*, Super-Enhancers in the Control of Cell Identity and Disease. *Cell*. **155**, 934–947 (2013).
155. J. R. Raab, R. T. Kamakaka, Insulators and promoters: Closer than we think. *Nat. Rev. Genet.* **11**, 439–446 (2010).
156. S. Holwerda, W. de Laat, Chromatin loops, gene positioning, and gene expression. *Front. Genet.* **3**, 1–13 (2012).
157. C. H. Waddington, The epigenotype. 1942. *Int. J. Epidemiol.* **41**, 10–13 (2012).
158. R. Holliday, The Inheritance of Epigenetic Defects. *Science (80-. )*. **238**, 163–170 (1987).
159. A. D. Riggs, G. P. Pfeifer, X-chromosome inactivation and cell memory. *Trends Genet.* **8**, 169–174 (1992).
160. N. C. Parashar, G. Parashar, H. Nayyar, R. Sandhir, N6-adenine DNA methylation

- demystified in eukaryotic genome: From biology to pathology. *Biochimie*. **144**, 56–62 (2018).
161. A. P. Bird, DNA methylation and the frequency of CpG in animal DNA. *Nucleic Acids Res.* **8**, 1499–1504 (1980).
  162. W. Xie *et al.*, Base-resolution analyses of sequence and parent-of-origin dependent DNA methylation in the mouse genome. *Cell*. **148**, 816–831 (2012).
  163. R. Lister *et al.*, Global epigenomic reconfiguration during mammalian brain development. *Science (80-. )*. **341**, 1237905 (2013).
  164. S. Saxonov, P. Berg, D. L. Brutlag, A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. *PNAS*. **103**, 1412–1417 (2006).
  165. T. H. Bestor, G. Gundersen, A. B. Kolstø, H. Prydz, CpG islands in mammalian gene promoters are inherently resistant to de novo methylation. *Genet. Anal. Biomol. Eng.* **9**, 48–53 (1992).
  166. N. R. Rose, R. J. Klose, Understanding the relationship between DNA methylation and histone lysine methylation. *Biochim. Biophys. Acta - Gene Regul. Mech.* **1839**, 1362–1372 (2014).
  167. Y. Liu, E. J. Oakeley, L. Sun, J. P. Jost, Multiple domains are involved in the targeting of the mouse DNA methyltransferase to the DNA replication foci. *Nucleic Acids Res.* **26**, 1038–1045 (1998).
  168. M. Okano, D. W. Bell, D. A. Haber, E. Li, DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell*. **99**, 247–257 (1999).
  169. A. Dhayalan *et al.*, The Dnmt3a PWWP domain reads histone 3 lysine 36 trimethylation and guides DNA methylation. *J. Biol. Chem.* **285**, 26114–26120 (2010).
  170. S. Chantalat *et al.*, Histone H3 trimethylation at lysine 36 is associated with constitutive and facultative heterochromatin. *Genome Res.* **21**, 1426–1437 (2011).
  171. M. Fatemi, A. Hermann, H. Gowher, A. Jeltsch, Dnmt3a and Dnmt1 functionally cooperate during de novo methylation of DNA. *Eur. J. Biochem.* **269**, 4981–4984 (2002).
  172. J. Arand *et al.*, In vivo control of CpG and non-CpG DNA methylation by DNA methyltransferases. *PLoS Genet.* **8** (2012), doi:10.1371/journal.pgen.1002750.
  173. G. Liang *et al.*, Cooperativity between DNA Methyltransferases in the Maintenance Methylation of Repetitive Elements. *Mol. Cell. Biol.* **22**, 480–491 (2002).
  174. R. J. Klose, A. P. Bird, Genomic DNA methylation: The mark and its mediators. *Trends Biochem. Sci.* **31**, 89–97 (2006).
  175. E. Hervouet, P. Peixoto, R. Delage-Mourroux, M. Boyer-Guittaut, P. F. Cartron, Specific or not specific recruitment of DNMTs for DNA methylation, an epigenetic dilemma. *Clin. Epigenetics*. **10**, 1–18 (2018).
  176. A. C. Bell, G. Felsenfeld, Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene. *Nature*. **405**, 482–485 (2000).
  177. B. Hendrich, A. Bird, Identification and Characterization of a Family of Mammalian Methyl-CpG Binding Proteins. *Mol. Cell. Biol.* **18**, 6538–6547 (1998).
  178. C. G. Spruijt *et al.*, Dynamic readers for 5-(Hydroxy)methylcytosine and its oxidized derivatives. *Cell*. **152**, 1146–1159 (2013).
  179. A. Prokhortchouk *et al.*, The p120 catenin partner Kaiso is a DNA methylation-dependent transcriptional repressor. *Genes Dev.* **15**, 1613–1618 (2001).
  180. S. Hu *et al.*, DNA methylation presents distinct binding sites for human transcription factors. *Elife*. **2013**, 1–16 (2013).
  181. H. Zhu, G. Wang, J. Qian, Transcription factors as readers and effectors of DNA methylation. *Nat. Rev. Genet.* **17**, 551–565 (2016).
  182. T. M. Geiman *et al.*, DNMT3B interacts with hSNF2H chromatin remodeling enzyme, HDACs 1 and 2, and components of the histone methylation system. *Biochem. Biophys. Res. Commun.* **318**, 544–555 (2004).
  183. A. Meissner *et al.*, Genome-scale DNA methylation maps of pluripotent and differentiated cells. *Nature*. **454**, 766–770 (2008).
  184. J. A. Yoder, C. P. Walsh, T. H. Bestor, Cytosine methylation and the ecology of intragenomic parasites. *Trends Genet.* **13**, 335–340 (1997).
  185. P. A. Jones, The DNA methylation paradox. *Trends Genet.* **15**, 34–37 (1999).
  186. A. Teissandier, D. Bourc'his, Gene body DNA methylation conspires with H3K36me3 to preclude aberrant transcription. *EMBO J.* **36**, 1471–1473 (2017).
  187. F. Neri *et al.*, Intragenic DNA methylation prevents spurious transcription initiation.



- Nature*. **543**, 72–77 (2017).
188. S. Shukla *et al.*, CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing. *Nature*. **479**, 74–79 (2011).
189. M. Tahiliani *et al.*, Conversion of 5-Methylcytosine to 5-Hydroxymethylcytosine in Mammalian DNA by MLL Partner TET1. *Science (80-. )*. **324**, 930–936 (2009).
190. S. Ito *et al.*, Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science (80-. )*. **333**, 1300–1303 (2011).
191. H. Wu, Y. Zhang, Reversing DNA methylation: Mechanisms, genomics, and biological functions. *Cell*. **156**, 45–68 (2014).
192. H. Huang, B. R. Sabari, B. A. Garcia, C. D. Allis, Y. Zhao, SnapShot : Histone Modifications. *Cell*. **159**, 458–458.e1 (2014).
193. T. Jenuwein, C. D. Allis, Translating the Histone Code. **293**, 1074–1081 (2001).
194. T. Kouzarides, Chromatin Modifications and Their Function. *Cell*. **128**, 693–705 (2007).
195. T. Agalioti, G. Chen, D. Thanos, Deciphering the transcriptional histone acetylation code for a human gene. *Cell*. **111**, 381–392 (2002).
196. N. D. Heintzman *et al.*, Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat. Genet.* **39**, 311–318 (2007).
197. N. D. Heintzman *et al.*, Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature*. **459**, 108–112 (2009).
198. M. J. Carrozza, R. T. Utley, J. L. Workman, J. Côté, The diverse functions of histone acetyltransferase complexes. *Trends Genet.* **19**, 321–329 (2003).
199. E. Seto, M. Yoshida, Erasers of Histone Acetylation : The Histone Deacetylase Enzymes. *Cold Spring Harb. Perspect. Biol.* **6**, a018713 (2014).
200. Z. Wu, J. Connolly, K. K. Biggar, Beyond histones – the expanding roles of protein lysine methylation. *FEBS J.* **284**, 2732–2744 (2017).
201. Y. Matsumura *et al.*, H3K4/H3K9me3 Bivalent Chromatin Domains Targeted by Lineage-Specific DNA Methylation Pauses Adipocyte Differentiation. *Mol. Cell*. **60**, 584–596 (2015).
202. P. J. Rugg-Gunn, B. J. Cox, A. Ralston, J. Rossant, Distinct histone modifications in stem cell lines and tissue lineages from the early mouse embryo. *Proc. Natl. Acad. Sci.* **107**, 10783–10790 (2010).
203. B. E. Bernstein *et al.*, A Bivalent Chromatin Structure Marks Key Developmental Genes in Embryonic Stem Cells. *Cell*. **125**, 315–326 (2006).
204. A. Barski *et al.*, High-Resolution Profiling of Histone Methylations in the Human Genome. *Cell*. **129**, 823–837 (2007).
205. M. A. Moro-García, R. Alonso-Arias, C. López-Larrea, When aging reaches CD4+ T-cells: Phenotypic and functional changes. *Front. Immunol.* **4**, 1–12 (2013).
206. A. H. F. M. Peters *et al.*, Loss of the Suv39h histone methyltransferases impairs mammalian heterochromatin and genome stability. *Cell*. **107**, 323–337 (2001).
207. T. Matsui *et al.*, Proviral silencing in embryonic stem cells requires the histone methyltransferase ESET. *Nature*. **464**, 927–931 (2010).
208. M. K. Ayrappetov, O. Gursoy-Yuzugullu, C. Xu, Y. Xu, B. D. Price, DNA double-strand breaks promote methylation of histone H3 on lysine 9 and transient formation of repressive chromatin. *PNAS*. **111**, 9169–9174 (2014).
209. J. Nakayama, J. C. Rice, B. D. Strahl, C. D. Allis, S. I. S. Grewal, Role of Histone H3 Lysine 9 Methylation in Epigenetic Control of Heterochromatin Assembly. *Science (80-. )*. **292**, 110–113 (2001).
210. M. García-Cao, R. O’Sullivan, A. H. F. M. Peters, T. Jenuwein, M. A. Blasco, Epigenetic regulation of telomere length in mammalian cells by the Suv39h1 and Suv39h2 histone methyltransferases. *Nat. Genet.* **36**, 94–99 (2004).
211. D. C. Schultz, K. Ayyanathan, D. Negorev, G. G. Maul, F. J. Rauscher III, SETDB1: a novel KAP-1-associated histone H3, lysine 9-specific methyltransferase that contributes to HP1-mediated silencing of euchromatic genes by KRAB zinc-finger proteins. *Genes Dev.* **16**, 919–932 (2002).
212. R. S. Allan *et al.*, An epigenetic silencing pathway controlling T helper 2 cell lineage commitment. *Nature*. **487**, 249–253 (2012).
213. S. Koide *et al.*, Setdb1 maintains hematopoietic stem and progenitor cells by restricting the ectopic activation of nonhematopoietic genes. *Blood*. **128**, 638–649 (2016).
214. J. Chen *et al.*, H3K9 methylation is a barrier during somatic cell reprogramming into iPSCs. *Nat. Genet.* **45**, 34–42 (2013).
215. G. E. Zentner, P. J. Tesar, P. C. Scacheri, Epigenetic signatures distinguish multiple classes of enhancers with distinct cellular

- functions. *Genome Res.* **21**, 1273–1283 (2011).
216. Y. Zhu, D. van Essen, S. Sacconi, Cell-Type-Specific Control of Enhancer Activity by H3K9 Trimethylation. *Mol. Cell.* **46**, 408–423 (2012).
  217. C. Stocking, C. A. Kozak, Murine endogenous retroviruses. *Cell. Mol. Life Sci.* **65**, 3383–3398 (2008).
  218. M. M. Karimi *et al.*, DNA Methylation and SETDB1/H3K9me3 Regulate Predominantly Distinct Sets of Genes, Retroelements, and Chimeric Transcripts in mESCs. *Cell Stem Cell.* **8**, 676–687 (2011).
  219. J. Sharif *et al.*, Activation of Endogenous Retroviruses in Dnmt1-/- ESCs Involves Disruption of SETDB1-Mediated Repression by NP95 Binding to Hemimethylated DNA. *Cell Stem Cell.* **19**, 81–94 (2016).
  220. T. S. Mikkelsen *et al.*, Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature.* **448**, 553–560 (2007).
  221. H. M. Rowe *et al.*, KAP1 controls endogenous retroviruses in embryonic stem cells. *Nature.* **463**, 237–240 (2010).
  222. R. Rebollo *et al.*, Retrotransposon-induced heterochromatin spreading in the mouse revealed by insertional polymorphisms. *PLoS Genet.* **7**, e1002301 (2011).
  223. H. M. Rowe *et al.*, TRIM28 repression of retrotransposon-based enhancers is necessary to preserve transcriptional dynamics in embryonic stem cells. *Genome Res.* **23**, 452–461 (2013).
  224. A. Bulut-Karslioglu *et al.*, Suv39h-Dependent H3K9me3 Marks Intact Retrotransposons and Silences LINE Elements in Mouse Embryonic Stem Cells. *Mol. Cell.* **55**, 277–290 (2014).
  225. P. L. Collins, K. E. Kyle, T. Egawa, Y. Shinkai, E. M. Oltz, The histone methyltransferase SETDB1 represses endogenous and exogenous retroviruses in B lymphocytes. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 8367–8372 (2015).
  226. A. Pasquarella *et al.*, Retrotransposon derepression leads to activation of the unfolded protein response and apoptosis in pro-B cells. (2016; <http://www.ncbi.nlm.nih.gov/pubmed/27013243>), vol. 143.
  227. S. L. Tan *et al.*, Essential roles of the histone methyltransferase ESET in the epigenetic control of neural progenitor cells during development. *Development.* **139**, 3806–3816 (2012).
  228. S. Liu *et al.*, Setdb1 is required for germline development and silencing of H3K9me3-marked endogenous retroviruses in primordial germ cells. *Genes Dev.* **28**, 2041–2055 (2014).
  229. C. R. Vakoc, S. A. Mandat, B. A. Olenchok, G. A. Blobel, Histone H3 Lysine 9 Methylation and HP1 $\gamma$  Are Associated with Transcription Elongation through Mammalian Chromatin. *Mol. Cell.* **19**, 381–391 (2005).
  230. V. Saint-André, E. Batsché, C. Rachez, C. Muchardt, Histone H3 lysine 9 trimethylation and HP1 $\gamma$  favor inclusion of alternative exons. *Nat. Struct. Mol. Biol.* **18**, 337–44 (2011).
  231. C. Mozzetta, E. Boyarchuk, J. Pontis, S. Ait-Si-Ali, Sound of silence: the properties and functions of repressive Lys methyltransferases. *Nat. Rev. Mol. Cell Biol.* **16**, 499–513 (2015).
  232. A. Loyola *et al.*, The HP1 $\alpha$ -CAF1-SetDB1-containing complex provides H3K9me1 for Suv39-mediated K9me3 in pericentric heterochromatin. *EMBO Rep.* **10**, 769–75 (2009).
  233. A. Loyola, T. Bonaldi, D. Roche, A. Imhof, G. Almouzni, PTMs on H3 Variants before Chromatin Assembly Potentiate Their Final Epigenetic State. *Mol. Cell.* **24**, 309–316 (2006).
  234. S. Beyer *et al.*, Canonical Wnt signalling regulates nuclear export of Setdb1 during skeletal muscle terminal differentiation. *Cell Discov.* **2** (2016), doi:10.1038/celldisc.2016.37.
  235. S. Cho, J. S. Park, Y. K. Kang, AGO2 and SETDB1 cooperate in promoter-targeted transcriptional silencing of the androgen receptor gene. *Nucleic Acids Res.* **42**, 13545–13556 (2014).
  236. M. Lachner, D. O'Carroll, S. Rea, K. Mechtler, T. Jenuwein, Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. *Nature.* **410**, 116–120 (2001).
  237. A. J. Bannister *et al.*, Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. *Nature.* **410**, 120–124 (2001).
  238. C. Muchardt *et al.*, Coordinated methyl and RNA binding is required for heterochromatin localization of mammalian HP1. *EMBO Rep.* **3**, 975–981 (2002).
  239. A. Nottke, M. P. Colaiacovo, Y. Shi, Developmental roles of the histone lysine demethylases. *Development.* **136**, 879–889 (2009).
  240. L. Wu *et al.*, Histone Demethylases KDM4A

- and KDM4C Regulate Differentiation of Embryonic Stem Cells to Endothelial Cells. *Stem Cell Reports*. **5**, 10–21 (2015).
241. Y. Satoh *et al.*, The Satb1 Protein Directs Hematopoietic Stem Cell Differentiation toward Lymphoid Lineages. *Immunity*. **38**, 1105–1115 (2013).
  242. P. P. Lee *et al.*, A critical role for Dnmt1 and DNA methylation in T cell development, function, and survival. *Immunity*. **15**, 763–74 (2001).
  243. O. M. Dovey *et al.*, Histone deacetylase 1 and 2 are essential for normal T-cell development and genomic stability in mice. *Blood*. **121**, 1335–1344 (2013).
  244. F. J. Martin *et al.*, KMT1E-mediated chromatin modifications at the FcγRIIb promoter regulate thymocyte development. *Genes Immun*. **16**, 162–169 (2015).
  245. S. Takikita *et al.*, A Histone Methyltransferase ESET Is Critical for T Cell Development. *J. Immunol.* (2016).
  246. B. Lehnertz *et al.*, Activating and inhibitory functions for the histone lysine methyltransferase G9a in T helper cell differentiation and function. *J. Exp. Med.* **207**, 915–22 (2010).
  247. B. Hao *et al.*, An anti-silencer- and SATB1-dependent chromatin hub regulates *Rag1* and *Rag2* gene expression during thymocyte development. *J. Exp. Med.* **212**, 809–824 (2015).
  248. R. Su *et al.*, Dynamic assembly of silent chromatin during thymocyte maturation. *Nat. Genet.* **36**, 502–506 (2004).
  249. X.-F. Zhou *et al.*, TRIM28 mediates chromatin modifications at the TCR enhancer and regulates the development of T and natural killer T cells. *Proc. Natl. Acad. Sci.* **109**, 20083–20088 (2012).
  250. F. R. Santoni de Sio *et al.*, KAP1 regulates gene networks controlling T-cell development and responsiveness. *FASEB J.* **26**, 4561–4575 (2012).
  251. I. Taniuchi, M. J. Sunshine, R. Festenstein, D. R. Littman, Evidence for distinct CD4 silencer functions at different stages of thymocyte differentiation. *Mol. Cell.* **10**, 1083–1096 (2002).
  252. M. Sellars *et al.*, Regulation of DNA methylation dictates Cd4 expression during the development of helper and cytotoxic T cell lineages. *Nat. Immunol.* **16**, 746–754 (2015).
  253. S. Delaire, Y. H. Huang, S. W. Chan, E. a Robey, Dynamic repositioning of CD4 and CD8 genes during T cell development. *J. Exp. Med.* **200**, 1427–35 (2004).
  254. Y. Kitagawa *et al.*, Guidance of regulatory T cell development by Satb1-dependent super-enhancer establishment. *Nat. Immunol.* **18**, 173–183 (2017).
  255. B. Liu, S. Tahk, K. M. Yee, G. Fan, K. Shuai, The Ligase PIAS1 Restricts Natural Regulatory T Cell Differentiation by Epigenetic. *Science (80-. )*. **330**, 521–525 (2010).
  256. G. Wei *et al.*, Global Mapping of H3K4me3 and H3K27me3 Reveals Specificity and Plasticity in Lineage Fate Determination of Differentiating CD4+ T Cells. *Immunity*. **30**, 155–167 (2009).
  257. M. Ciofani *et al.*, A validated regulatory network for Th17 cell specification. *Cell*. **151**, 289–303 (2012).
  258. R. D. Hawkins *et al.*, Global Chromatin State Analysis Reveals Lineage-Specific Enhancers during the Initiation of Human T helper 1 and T helper 2 Cell Polarization. *Immunity*. **38**, 1271–1284 (2013).
  259. F. Zhang, M. Boothby, T helper type 1-specific Brg1 recruitment and remodeling of nucleosomes positioned at the IFN-γ promoter are Stat4 dependent. *J. Exp. Med.* **203**, 1493–1505 (2006).
  260. A. L. Wurster, M. J. Pazin, BRG1-Mediated Chromatin Remodeling Regulates Differentiation and Gene Expression of T Helper Cells. *Mol. Cell. Biol.* **28**, 7274–7285 (2008).
  261. O. Hakim *et al.*, Spatial congregation of STAT binding directs selective nuclear architecture during T-cell functional differentiation. *Genome Res.* **23**, 462–472 (2013).
  262. S. Cai, C. C. Lee, T. Kohwi-Shigematsu, SATB1 packages densely looped, transcriptionally active chromatin for coordinated expression of cytokine genes. *Nat. Genet.* **38**, 1278–1288 (2006).
  263. A. Onodera *et al.*, STAT6-mediated displacement of polycomb by trithorax complex establishes long-term maintenance of GATA3 expression in T helper type 2 cells. *J. Exp. Med.* **207**, 2493–2506 (2010).
  264. A. Wang *et al.*, Cutting Edge: Smad2 and Smad4 Regulate TGF-β – Mediated IL9 Gene Expression via EZH2 Displacement. *J. Immunol.* **191**, 4908–4912 (2013).
  265. K. Ichiyama *et al.*, The Methylcytosine Dioxygenase Tet2 Promotes DNA Demethylation and Activation of Cytokine Gene Expression in T Cells. *Immunity*. **42**, 613–626 (2015).



266. C. L. Williams et al., STAT4 and T-bet Are Required for the Plasticity of IFN-  $\gamma$  Expression across Th2 Ontogeny and Influence Changes in Ifng Promoter DNA Methylation. *J. Immunol.* **191**, 678–687 (2013).
267. K. W. Makar et al., Active recruitment of DNA methyltransferases regulates interleukin 4 in thymocytes and T cells. *Nat. Immunol.* **4**, 1183–1190 (2003).
268. K. W. Makar, C. B. Wilson, DNA methylation is a nonredundant repressor of the Th2 effector program. *J. Immunol.* **173**, 4402–6 (2004).
269. R. M. Thomas, C. J. Gamper, B. H. Ladle, J. D. Powell, A. D. Wells, De Novo DNA methylation is required to restrict T helper lineage plasticity. *J. Biol. Chem.* **287**, 22900–22909 (2012).
270. D. J. Tumes et al., The Polycomb protein Ezh2 regulates differentiation and plasticity of CD4+ T helper Type 1 and type 2 cells. *Immunity.* **39**, 819–832 (2013).
271. Y. Zhang et al., The polycomb repressive complex 2 governs life and death of peripheral T cells. *Blood.* **124**, 737–750 (2014).
272. M. Koyanagi et al., EZH2 and Histone 3 Trimethyl Lysine 27 Associated with Il4 and Il13 Gene Silencing in TH1 Cells. *J. Biol. Chem.* **280**, 31470–31477 (2005).
273. V. Ranzani et al., The long intergenic noncoding RNA landscape of human lymphocytes highlights the regulation of T cell differentiation by linc-MAF-4. *Nat. Immunol.* **16**, 318–325 (2015).
274. A. Arvey et al., Inflammation-induced repression of chromatin bound by the transcription factor Foxp3 in regulatory T cells. *Nat. Immunol.* **15**, 580–588 (2014).
275. M. DuPage et al., The chromatin-modifying enzyme Ezh2 is critical for the maintenance of regulatory T cell identity after activation. *Immunity.* **42**, 227–238 (2015).
276. S. L. Hewitt, F. A. High, S. L. Reiner, A. G. Fisher, M. Merkenschlager, Nuclear repositioning marks the selective exclusion of lineage-inappropriate transcription factor loci during T helper cell differentiation. *Eur. J. Immunol.* **34**, 3604–3613 (2004).
277. X. Xiao et al., The Costimulatory Receptor OX40 Inhibits Interleukin-17 Expression through Activation of Repressive Chromatin Remodeling Pathways. *Immunity.* **44**, 1271–1283 (2016).
278. C. Huang et al., Cutting Edge: A Novel, Human-Specific Interacting Protein Couples FOXP3 to a Chromatin-Remodeling Complex That Contains KAP1/TRIM28. *J. Immunol.* **190**, 4470–4473 (2013).
279. Y. Wakabayashi et al., Histone 3 Lysine 9 (H3K9) Methyltransferase Recruitment to the Interleukin-2 (IL-2) Promoter Is a Mechanism of Suppression of IL-2 Transcription by the Transforming Growth Factor-  $\beta$ -Smad Pathway. *J. Biol. Chem.* **286**, 35456–35465 (2011).
280. Y. Jiang et al., Epigenetic activation during T helper 17 cell differentiation is mediated by Tripartite motif containing 28. *Nat. Commun.* **9** (2018), doi:10.1038/s41467-018-03852-2.
281. J. O'Shea, W. E. Paul, Mechanisms underlying lineage commitment and plasticity of helper CD4 + T cells. *Science (80-. ).* **327** (2010), pp. 1098–1102.
282. C. B. Wilson, E. Rowell, M. Sekimata, Epigenetic control of T-helper-cell differentiation. *Nat. Rev. Immunol.* **9**, 91–105 (2009).
283. Y. Kanno, G. Vahedi, K. Hirahara, K. Singleton, J. J. O'Shea, Transcriptional and epigenetic control of T helper cell specification: molecular mechanisms underlying commitment and plasticity. *Annu. Rev. Immunol.* **30**, 707–31 (2012).
284. H. A. Young et al., Differentiation of the T helper phenotypes by analysis of the methylation state of the IFN-gamma gene. *J. Immunol.* **153**, 3603–3610 (1994).
285. J. J. O'Shea, R. Lahesmaa, G. Vahedi, A. Laurence, Y. Kanno, Genomic views of STAT function in CD4 + T helper cell differentiation. *Nat. Rev. Immunol.* **11**, 239–250 (2011).
286. S. Bilodeau, M. H. Kagey, G. M. Frampton, P. B. Rahl, R. A. Young, SetDB1 contributes to repression of genes encoding developmental regulators and maintenance of ES cell state. *Genes Dev.* **23**, 2484–2489 (2009).
287. J. Liu et al., Chromatin Landscape Defined by Repressive Histone Methylation during Oligodendrocyte Differentiation. *J. Neurosci.* **35**, 352–365 (2015).
288. J. S. Becker, D. Nicetto, K. S. Zaret, H3K9me3-Dependent Heterochromatin: Barrier to Cell Fate Changes. *Trends Genet.* **32**, 29–41 (2016).
289. A. H. F. M. Peters et al., Histone H3 lysine 9 methylation is an epigenetic imprint of facultative heterochromatin. *Nat. Genet.* **30**, 77–80 (2002).
290. E. B. Chuong, N. C. Elde, C. Feschotte, Regulatory activities of transposable

- elements: from conflicts to benefits. *Nat. Rev. Genet.* **18**, 71–86 (2017).
291. P. L. Collins, K. E. Kyle, T. Egawa, Y. Shinkai, E. M. Oltz, The histone methyltransferase SETDB1 represses endogenous and exogenous retroviruses in B lymphocytes. *Proc. Natl. Acad. Sci.* **112**, 8367–8372 (2015).
  292. K. B. Chiappinelli et al., Inhibiting DNA Methylation Causes an Interferon Response in Cancer via dsRNA Including Endogenous Retroviruses. *Cell*. **162**, 974–986 (2015).
  293. J. D. Milner, J. M. Ward, A. Keane-Myers, W. E. Paul, Lymphopenic mice reconstituted with limited repertoire T cells develop severe, multiorgan, Th2-associated inflammatory disease. *PNAS*. **104**, 576–581 (2007).
  294. E. B. Chuong, N. C. Elde, C. Feschotte, Regulatory evolution of innate immunity through co-option of endogenous retroviruses. *Science (80-. )*. **351**, 1083–1087 (2016).
  295. M. Imbeault, P.-Y. Helleboid, D. Trono, KRAB zinc-finger proteins contribute to the evolution of gene regulatory networks. *Nature*. **543**, 550–554 (2017).
  296. C. Trapnell, L. Pachter, S. L. Salzberg, TopHat: Discovering splice junctions with RNA-Seq. *Bioinformatics*. **25**, 1105–1111 (2009).
  297. T. Lee, S. Johnston, R. Young, Chromatin immunoprecipitation from *C. elegans* embryos. *Nat. Protoc.* **1**, 729–748 (2006).
  298. H. Li, R. Durbin, Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. **25**, 1754–1760 (2009).
  299. Y. Zhang et al., Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9** (2008), doi:10.1186/gb-2008-9-9-r137.
  300. A. T. L. Lun, G. K. Smyth, Cseq: A Bioconductor package for differential binding analysis of ChIP-seq data using sliding windows. *Nucleic Acids Res.* **44** (2015), doi:10.1093/nar/gkv1191.
  301. J. D. Buenrostro, B. Wu, H. Y. Chang, W. J. Greenleaf, *Curr. Protoc. Mol. Biol.*, in press, doi:10.1002/0471142727.mb2129s109.
  302. J. Göke et al., Dynamic transcription of distinct classes of endogenous retroviral elements marks specific populations of early human embryonic cells. *Cell Stem Cell*. **16**, 135–141 (2015).
  303. S. Anders, P. T. Pyl, W. Huber, HTSeq-A Python framework to work with high-throughput sequencing data. *Bioinformatics*. **31**, 166–169 (2015).
  304. A. R. Quinlan, I. M. Hall, BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics*. **26**, 841–842 (2010).
  305. A. Akalin, V. Franke, K. Vlahoviček, C. E. Mason, D. Schübeler, Genomation: A toolkit to summarize, annotate and visualize genomic intervals. *Bioinformatics*. **31**, 1127–1129 (2015).
  306. C. Y. McLean et al., GREAT improves functional interpretation of cis-regulatory regions. *Nat. Biotechnol.* **28**, 495–501 (2010).
  307. R. C. Mcleay, T. L. Bailey, Motif Enrichment Analysis: a unified framework and an evaluation on ChIP data Motif Enrichment Analysis: a unified framework and an evaluation on ChIP data (2010), doi:10.1186/1471-2105-11-165.
  308. G. Bourque et al., Evolution of the mammalian transcription factor binding repertoire via transposable elements. *Genome Res.*, 1752–1762 (2008).
  309. V. Sundaram et al., Widespread contribution of transposable elements to the innovation of gene regulatory networks. *Genome Res.* **24**, 1963–1976 (2014).
  310. V. Sundaram et al., Functional cis-regulatory modules encoded by mouse-specific endogenous retrovirus. *Nat. Commun.* **8**, 14550 (2017).
  311. M. Xie et al., DNA hypomethylation within specific transposable element families associates with tissue-specific enhancer landscape. *Nat. Genet.* **45**, 836–841 (2013).
  312. E. B. Chuong, M. A. K. Rumi, M. J. Soares, J. C. Baker, Endogenous retroviruses function as species-specific enhancer elements in the placenta. *Nat. Genet.* **45**, 325–329 (2013).
  313. T. L. Cuellar et al., --Silencing of retrotransposons by SETDB1 inhibits the interferon response in acute myeloid leukemia--. *J. Cell Biol.* **216**, 3535–3549 (2017).
  314. G. Wolf et al., The KRAB zinc finger protein ZFP809 is required to initiate epigenetic silencing of endogenous retroviruses. *Genes Dev.* **29**, 538–554 (2015).
  315. L. De Monte et al., Intratumor T helper type 2 cell infiltrate correlates with cancer-associated fibroblast thymic stromal lymphopoietin production and reduced survival in pancreatic cancer. *J. Exp. Med.* **208**, 469–78 (2011).
  316. H. Ubukata et al., Evaluations of interferon- $\gamma$ /interleukin-4 ratio and neutrophil/lymphocyte ratio as prognostic indicators in gastric cancer patients. *J. Surg. Oncol.* **102**, 742–747 (2010).

317. M. L. Suvà, N. Riggi, B. E. Bernstein, Epigenetic Reprogramming in Cancer. *Nat. Med.* **1567**, 179–92 (2013).
318. C. J. Ceol et al., The histone methyltransferase SETDB1 is recurrently amplified in melanoma and accelerates its onset. *Nature*. **471**, 513–517 (2012).

# RESUME

Les lymphocytes T CD4 naïfs sont des cellules plastiques, capables de moduler finement leur programmation selon les signaux environnementaux qu'ils intègrent. Ils adaptent ainsi leur phénotype et leur fonction au type de danger lors d'une infection par un agent pathogène intracellulaire par exemple, ils acquièrent un phénotype Th1 sous l'influence de médiateurs solubles tels que l'IL-12 et l'IFN- $\gamma$ . Ces signaux mobilisent un set restreint de facteurs de transcription, coordonné par Tbet, qui programment la cellule afin qu'elle induise l'élimination du danger par des mécanismes impliquant une production massive d'IFN- $\gamma$ . En réponse à des allergènes ou à des parasites extracellulaires, les lymphocytes T peuvent aussi acquérir un phénotype Th2, caractérisé par l'expression du facteur de transcription Gata-3 et par la production d'IL-4, d'IL-5 et d'IL-13. Afin de garantir la stabilité des lignages, ces processus de différenciation peuvent s'accompagner d'une perte de potentialité. Contrairement aux cellules T naïves, les cellules Th1 sont par exemple incapables d'allumer le programme d'expression génique Th2 en présence d'IL-4, et les lymphocytes Th2 verrouillent le programme Th1. Si nous savons aujourd'hui que l'acquisition des fonctions effectrices, comme l'équilibre entre détermination cellulaire et plasticité, sont régulés par des mécanismes épigénétiques, la plupart des acteurs moléculaires qui contrôlent la programmation des lymphocytes T au niveau de la chromatine reste encore à identifier. Durant ma thèse, j'ai étudié le rôle de la lysine méthyltransférase SETDB1, qui catalyse la di- ou tri-méthylation de la lysine 9 de l'histone 3 (H3K9me3), dans la différenciation des lymphocytes T CD4. Il avait déjà été proposé qu'H3K9me3 ait un impact sur la programmation de ces cellules en réponse aux signaux de l'environnement, mais personne n'avait encore étudié le rôle de SETDB1 dans ces processus lorsque j'ai commencé ma thèse. A l'aide d'une lignée murine déficiente pour SETDB1 spécifiquement dans les lymphocytes T, nous avons montré *in vitro* et *in vivo* que la balance Th1/Th2 est fortement augmentée en l'absence de l'enzyme, et que cette dérégulation résulte d'une perte de répression du réseau génique Th1. Au niveau mécanistique, nos analyses moléculaires ont mis en évidence que l'absence de SETDB1 dans les lymphocytes Th2 s'accompagne d'une perte d'H3K9me3 sur un réseau restreint de rétrovirus endogènes (ERV) qui se comportent comme des modules cis-régulateurs de gènes clefs du programme Th1. Nous avons donc mis en évidence que SETDB1 garantit l'intégrité des lignages T helpers en réprimant, via la déposition d'H3K9me3, un set restreint d'ERV exaptés en enhanceurs de gènes Th1.